

494 TrpGluValLeuAlaArgHisGlySerIleTyrValAsnGluGlnPheThrLysLeuVal 513
1381 TGGGAGTTTGGCGAGGACCGCTCCAGATCTGGTCAATGAGGAGACCAAGCTGGTG 1440
514 TyrPheGlnGlyThrLysAspThrProLeuGluHisHisLeuTyrValValSerTyrGlu 533
1441 TACTTCCAGGCGACCAAGACACGCGCTGGAGACCACTCTACGTGGTTCAGCTATGAG 1500
534 AlaAlaGlyGluIleValArgLeuThrThrProGlyPheSerHisSerCysSerMetSer 553
1501 GCGGCGCGGAGATCGCTACCGCTCCAGCGCGGTTCCTCCATAGCTGCTCCATGAGC 1560
554 GlnAsnPheAspMetPheValSerHisTyrSerSerValSerThrProCysValHis 573
1561 CAGAACTTCGACATGCTCTGACCACTACAGCAGCTGAGCAGCGCGCTGGTGCAC 1620
574 ValTyrLysLeuSerGlyProAspAspProLeuHisLysGlnProArgPheTrpAla 593
1621 GTCTACAAGCTGAGCGCGCCGAGCAGACCCCTGCACAAGCAGCGCGCTTCTGGGCT 1680
594 SerMetMetGluAlaAlaSerCysProProAspTyrValProGluIlePheHisPhe 613
1681 AGCATGATGAGGACCAACAGCTGCCCCCGGATATGTTCTCCAGAGATCTTCCATTTC 1740
614 HisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnPro 633
1741 CACACGCGCTCGATGCTGGCTCTACGCGATGATCTACAAGCCCGCAGCTGCGACCA 1800
634 GlyLysLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAsn 653
1801 GGGAGAAGCACCCACCGCTCTTTTGTATATGAGGCGCCCGCAGCTGCGTGAAT 1860
654 AsnSerPheLysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAla 673
1861 AACTCTTCAAGGCGATCAAGTACTTGGCGCTCAACACACTGGCGCTCCCTGGGCTACGCC 1920
674 ValValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeu 693
1921 GTGGTGTGATTCAGCGCAGGCGCTCTGTCAGCGAGGCTTCGGTTCGAGGGCGCTG 1980
694 LysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAla 713
1981 AAAACCAATGGCCAGGCTGGAGATCGAGGACCGAGGCGGCTGCGAGTTCGTGGCC 2040
714 GlnLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTyrSerTyrGlyGly 733
2041 GAGAAGTATGGCTTCATCGACCTGAGCGAGTTGCCATCCATGGCTGCTACGGGGC 2100
734 PheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaGly 753
2101 TTCTCTCGCTATGGGCTTAATCCAAAGCCCGAGGTGTCAAGTGGCCATCGCGGGT 2160
754 AlaProValThrValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspVal 773
2161 GCCCGGTACCGCTGATGGCTACGACAGGCTACAGTGGCTGCTACATGGAGCTC 2220
774 ProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeu 793
2221 CCTGAGAACAACCAAGCAGCGCTATGAGCGGGTTCCTGGCGCTGCGAGTGGAGAAGCTG 2280
794 ProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhe 813
2281 CCCAATGAGCCCAACCGCTGCTTATCTCCACGCGCTCTCCGAGCAAAAGCTGCATTT 2340
814 PheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGln 833
2341 TTCACACAAACTTCCTCGTCTCCCACTGATCCGAGCAGGGAACCTTACACGCTCCAG 2400
834 IleTyrProAsnGluArgHisSerIleArgCysProGluSerGlyGluHisTyrGluVal 853
2401 ATCTACCCCAACGAGACACAGATTCGCTGCCCGAGTCCGCGGCGAGCATTATGAAGTC 2460

854 ThrLeuLeuHisPheLeuGlnGluTyrLeu 863
2461 ACGTACTGCACTTCTACAGGATACCTC 2490
RESULT 8
AAI57896
ID AAI57896 standard; cDNA; 2801 BP.
XX AAI57896;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 99.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0489725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620315.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;
WPI: 2001-442253/47.
P-PSDB; AAM38740.
XX Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -
XX Claim 1; SEQ ID NO 99; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
the encoded polypeptides (AAM38642-AAM42213) with nootropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: Immune system suppression,
activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemia and
C.N.S. diseases.
XX Sequence 2801 BP; 586 A; 889 C; 801 G; 525 T; 0 other;
Alignment Scores: 0 Length: 2801
Pred. No.: 0

Score: 4352.00 Matches: 827
Percent Similarity: 94.32% Conservative: 3
Best Local Similarity: 93.98% Mismatches: 5
Query Match: 93.67% Indels: 45
DB: 22 Gaps: 5

US-09-976-674-3 (1-863) x AAI57896 (1-2801)

QY 1 MetalThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaAlaThrAspPro 20
~
Db 245 ATGGCCACACCGGACCCCAACAGCGCCGACCGAGCGGACGCCACACATGACCG 304
QY 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleHisGly 40
~
Db 305 GCCGCCGGTCCAGTGCAGAGCACTCGTGGAGGGCTCCGGAGCATCATCCACGGC 364
QY 41 SerArgLysTrpSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
~
Db 365 AGCCCAAGTACTCGGCCCTCATTTGCAACAGGCGCCCGACGACTTCAGTTTGTGCG 424
61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTrpLeuGlyMetProTrpGly 80
~
Db 425 AAGACGGATGACTTGGGCCACCTCCACCGCCCTCTACTAGCTGGGTATGCATATGC 484
QY 81 SerArgGluAsnSerLeuLysSerGluIleProLysLysValArgLysGluAlaLeu 100
~
Db 485 AGCCGAGAACTCCCTCTCTACTCTGAGATCCCAAGAGTCCGGAAGAGGCTCTG 544
QY 101 LeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisGlyVal 120
~
Db 545 CTGCTCTCTGCGAAGCAGATGCTGGATCATTTCCAGGCCACGCCACCATGGGTC 604
QY 121 TyrSerArgGluGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr 140
~
Db 605 TACTCTCGGGAGGAGAGTCTGTAGGGAGCGGAACGCTGGGGCTCTTCGGCATCAC 664
QY 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
~
Db 665 TCCTACGACTTCCACAGCAGAGTGGCTCTCTCTTCCAGGCCACGACGCTCTTC 724
QY 161 HisCysArgAspGlyLysAsnGlyPheMet-----ValSer 173
~
Db 725 CACTGCGCGAGCGGCGCAAGACGGCTTCATGGTGAGCCCTGGCCCTGTGTGTCC 784
QY 174 ProMetLysProLeuGluIleLysThrGlnCysSerGlyProArgMetAspProLysIle 193
~
Db 785 CCTATGAACCGCTGGAATCAAGACCCAGTCTCAGGGCCCGGATGACCCCAAAATC 844
194 CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTrpValAlaAsn 213
~
Db 845 TGCCCTGCGACCCCTGCTCTCTCTCTCATCAATAACAGCGACCTGTGGTGGCCAC 904
QY 214 IleGluThrGlyGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 233
~
Db 905 ATCGACACAGGCGAGGCGGGCTGAGCTTCTGCCACCAAGGTTTATCAATGCTCTG 964
QY 234 AspAspProLysSerAlaGlyValAlaThrPheValIleGlnGluPheAspArgPhe 253
~
Db 965 GATGACCCCAAGTCTCGGGGTGGGCCACTTCGTCTACAGGAAGTTCGACCGCTTC 1024
QY 254 ThrGlyTyrTrpCysProThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeu 273
~
Db 1025 ACTGGGTACTGGTGGTGGCCACAGCTCTCTGGAGAGGTTCAGAGGCTTCAGAGCTG 1084
QY 274 ArgIleLeuTyrGluGluValAspGluSerGluValGlnValIleHisValProSerPro 293
~
Db 1085 CGAATCTGTATGAGGAAGTCGATGATCGGAGTGGAGTCAATTCACGCTCCCTCTCT 1144
QY 294 AlaLeuGluGluArgLysThrAspSerTyrArgTyrProArgThrGlySerLysAsnPro 313
~
Db 1145 GCGCTAGAGAAGGAGAGCGGACTCGTATCGGTATCCCGGAGGACGACGACAGAAATCCC 1204
QY 314 LysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThr 333

Db 1205 AAGATTGCTTGAACCTGGCTGAGTCCAGACTGACGCCAGGCGAAGATCGTCTCGACC 1264
QY 334 GlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysValGluTyrIleAla 353
~
Db 1265 CAGGAGAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTTCCCGAAGGTGAGTACATCCGC 1324
QY 354 ArgAlaGlyTrpThrArgAspGlyLysTyrAlaTrpAlaMetPheLeuAspArgProGln 373
~
Db 1325 AGGCGCGGTGGACCGGGATGGCAATATACCTCGGGCATATGTTCTCTGGACCGGCCAG 1384
QY 374 GlnTrpLeuGlnLeuValLeuProProAlaLeuPheIleProSerThrGluAsnGlu 393
~
Db 1385 CAGTGGCTTCCAGCTCGTCTCTCCCGGCCCTGTTCTATCCCGAGCAGAGANTGAG 1444
QY 394 GluGln-ArgLeuAlaSerAlaArgAlaValProArgAsnValGlnProTrpValVal 413
~
Db 1445 GAGCAGCGGGCTAGCTCTGCCAGAGCTGTCCCGAGGAATGTCAGCGCGTATGTGTGTAC 1504
QY 413 rGluGluValThrAsnValTrpIleAsnValHisAspIlePheTyrProPheProGlnSe 433
~
Db 1505 GGAGGAGTCAACCAACGCTCTGGATCAATGTTCTATGACATCTTCTATCCCTTCCCAATC 1564
QY 433 rGluGlyGluAspGluLeuCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysH 453
~
Db 1565 AGAGGAGAGGAGCAGCTCTCTCTCCCGGCCAATGAATGCAAGACCGGCTTCTGCA 1624
QY 453 sLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGluProPheSe 473
~
Db 1625 TTTGTACAAAGTCACCGCGCTTTAAATCCAGGGCTACGATTGGAGTGAGCCTTCAG 1684
QY 473 rProGlyGluAspGluPheLysCysProIleLysGluIleAlaLeuThrSerGlyG 493
~
Db 1685 CCCCAGGGAAGGT-----GAGCAGACGCTGACGAATGCT-- 1718
QY 493 uTrpGluValLeuAlaArgHisGlySerLysIleTrpValAsnGluLysLeuVal 513
~
Db 1719 -----ATCTGGGTCAATGAGGAGACCAAGCTGGT 1747
QY 513 lTyrPheGlnGlyThrLysAspThrProLeuGluHisHisLeuTyrValValSerTrpG 533
~
Db 1748 GTACTTCCAGGCGACCAAGGACACCGCTGGAGCACCCTCTACGTGTGCTCAGTATGA 1807
QY 533 uAlaAlaGlyIleValArgLeuThrThrProGlyPheSerHisSerCysSerMetSe 553
~
Db 1808 GCGGCGGCGAGATCGTACGCTCACCAGCCCGCTCTCCCATAGTGTCTCCTCAGT 1867
QY 553 rGlnAsnPheAspMetPheValSerHisTyrSerValSerThrProProCysValH 573
~
Db 1868 CCAGAACTTCGACATGTCGTACGCACTACAGAGCGTGAGCAGCCGCTCGGTGCA 1927
QY 573 sValTyrLysLeuSerGlyProAspAspProLeuHisLysGlnProArgPheTrpAl 593
~
Db 1928 CGTCTACAGCTGAGCGGCGCGGAGCAGACCCCTGCAACAGCAGCCCGCTTCTGGGC 1987
QY 593 aSerMetMetGluAlaAlaSerCysProProAspTrpValProProGluIlePheHisPh 613
~
Db 1988 TAGCATGATGGAGGAGCGC-----AAGATCTTCATTT 2020
QY 613 eHisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnPr 633
~
Db 2021 CCACAGCGCTCGGATGTCGGCTCTACGGCATGATCTACAGCCCGCCAGCCTTCAGGC 2080
QY 633 oGlyLysLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAs 653
~
Db 2081 AGGGAAGAAGCACCCACCGCTCTCTTTGTATATGAGGCGCCCGGAGTGCAGTGGTGAA 2140
QY 653 nasnSerPheLysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTrAl 673
~
Db 2141 TAACCTCTTCAAGGCATCAAGTACTTGGGTCAACACACTGGCCTCCCTGGGTATCC 2200
QY 673 aValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLe 693

Db 2201 CGGTGTTGATTGACGGAGGGGCTCTGTCAGCAGAGGCTTCGGTTCGAAGGGCCCT 2260
 QY 693 uTysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAl 713
 Db 2261 GAAAAACCAATGGCCAGATGGAGATCGAGACCAAGTGGAGGGCTGAGTTCGTGGC 2320
 QY 713 AGULYSTRYGLYPHLEIASPLEUSERARGVALAALLEHISGLYTPSERTYRGLVGI 733
 Db 2321 CCGAGATATGGCTTCATGACCTGAGCCGACAGTTCCTCCATCCAGGCTGGTCTTACGGGG 2380
 QY 733 yPheLusSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaGI 753
 Db 2381 CTTCCTGCTGCTATGAGGGGTAATCCACAGCCCGAGGTTCAGAGTGGCCATTCGGGG 2440
 QY 753 yAlaProValThrValITrpMetAlaTyrAspPheGlyTyrThrGluArgTyrMetAspVa 773
 Db 2441 TGGCCCGGTACCGCTCGATGGCTTACGACAGGGGTACACAGGGCTACATGACGT 2500
 QY 773 lProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLe 793
 b 2501 CCTGAGAACACACAGACGCTATGAGGGGGTTCGTGGCCCTGACGCTGAGAGAGCT 2560
 QY 793 uProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnValHisPh 813
 Db 2561 GCCCAATGACCCCAACCGCTTCTTATCTCCACGCTTCTGACCAAAAACGTCACCT 2620
 QY 813 ePheHisThrAspPheLeuValSerGlnLeuIleArgAlaGlyAspProTyrGlnLeu-- 832
 Db 2621 TTTCACACAAACTCTCTCGTCCCAACTGATCCGAGCAGGAAACCTTACCAAGCTCCA 2680
 QY 833 -----GlnIleTyrProAsnGluArgHisSerIleArgcy 844
 Db 2681 GGTGGCCCTGCTCTGCTCTCCCGCAGATCTACCCCAACAGAGACACATATTCGCTG 2740
 QY 844 sProGluSerGlyGlnHisTyrGluValIThrLeuLeuHisPheLeuGlnGluTyrLeu 863
 Db 2741 CCCGAGTGGGGCAGCACTATGAAAGTACGTCGTGCACTTCTACAGAAATACCTC 2798
 RESULT 9
 ABR83337
 ID ABR83337 standard; cDNA; 4076 BP.
 AC ABR83337;
 DT 12-AUG-2002 (first entry)
 XX
 DE cDNA encoding human DPRP-2 splice variant #5.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KW DPTIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 KW
 OS Homo sapiens.
 XX
 PN WO200231134-A2.
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;
 DR WPI: 2002-444178/47.
 DR P-PSDB; ABR61606.
 XX

PR New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PR the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PR viral infections, cancers, allergies, neurological disorders, or pain
 PS
 XX Disclosure: Page 90-91; 113pp; English.
 CC
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPTIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections,
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bullimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABR83322-ABR83343 encode human DPRP proteins.
 CC
 SQ Sequence 4076 BP; 879 A; 1276 C; 1143 G; 778 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 4076
 Score: 4337.00 Matches: 814
 Percent Similarity: 94.33% Conservative: 1
 Best Local Similarity: 94.21% Mismatches: 0
 Query Match: 93.35% Indels: 49
 DB: 24 Gaps: 1
 US-09-976-674-3 (1-863) x ABR83337 (1-4076)
 QY 1 MetaLarThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspAspPro 20
 Db 367 ATGGCCACACACCGGGAGCCCAAGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGG 426
 QY 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
 Db 427 GCCCGCCGCTTCCAGGTGACAGAACACCTCGTGGAGCGGCGCCGAGCATATCCACGGC 486
 QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
 Db 487 AGCCGACAGTACTCGGGCTCATTTGTCAACAGCGCCCGCCAGCTTCAGATTGTGAG 546
 QY 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
 Db 547 AAGACGATGAGTCTGGGCCCCACCTCCACCGCGCTCTACTACGGAATGCAATATGGC 606
 QY 81 SerArgLysAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
 Db 607 AGCCGACAGTACTCGGGCTCATTTGTCAACAGCGCCCGCCAGCTTCAGATTGTGAG 666
 QY 101 LeuLeuLusSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
 Db 667 CTGCTCTGCTCTGGAACAGATGCTGGATTCATTCAGAGCCACGCCCAACATGGGGGC 726
 QY 121 TyrSerArgGluGluGlnLeuLeuArgLysLysArgLeuGlyValPheGlyIleThr 140
 Db 727 TACTCTCGGAGAGAGAGAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 786
 QY 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
 Db 787 TCCTACGACTTCCACAGGAGAGTGGCTCTTCTTCCAGGACGACCAAGCCTCTTC 846
 QY 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
 Db 847 CACTCCCGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 906
 QY 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
 Db 907 AAGACCAAGTGTCTAGGGGCGCGGAGTGAACCCCAAAATCTGCCCTCCGACCTGCTTC 966


```

|||||
Db 653 TyGClYgLYProclnValGlnLeuValAsnAspArgPheLYgLYValLYsTYrPheArg 672
QY 2000 CTCACACACTGGCTCCCTGGGCTACGCGCTGTTGTATGAGCGAGCGGCTCTGT 2059
Db 673 LeuAsnThrLeuAlaSerLeuGlyTYrValValValValIleAspAsnArgLYserGys 692
QY 2060 CAGCGAGGCTGGTGGTCCAGAGGCGCCGAAAACCAATGGCCAGGTGAGATCGAG 2119
Db 693 HisArgGlyLeuLYsPheGlyGlyAlaPheLYsTYrLYsMetGlyGlnIleGlnLeuAsp 712
QY 2120 GACCAAGTGGAGGCGCTGACGTTCTGTGCGCGAGAGTATGCTCATGACCTGAGCCGA 2179
Db 713 AspGlnValGlnGlyLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeuAspArg 732
QY 2180 GTTGCCATCATGGCTGTGCTTACGGGCGCTCTCTGCTCATGAGGCGCTATCCACAG 2239
Db 733 ValGlyLeuHisLYsLYrPheTYrGlyGlyTYrLeuSerLeuMetAlaLeuMetGlnArg 752
QY 2240 CCCCAGGTGTTCAAGGTGACGTCGCGGCTGCGCGGCTGACGCTGATGAGCTACGAG 2239
Db 753 SerAspPheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTYrAsp 772
QY 2300 ACAGGCTACCTAGCGCTACATGAGAGCTCCCTGAGAACACACAGCGCTATGAGCG 2359
Db 773 ThrGlyTYrThrGlnArgTYrMetGlyHisProAspGlnAsnGlyGlnIleTYrTYrLeu 792
QY 2360 GGTTCCGTGGCCCTGACGCTGAGAGCTGCCCAATGACCCCAACCGCTGCTTCTCTC 2419
Db 793 GlySerValAlaMetGlnAlaGlyLYsPheProSerGlnProAsnArgLeuLeuLeu 812
QY 2420 CACGGCTCCCTGACGAGAAAGCTGACCTTTTTCACACAACTTCCTGCTCCCACTG 2479
Db 813 HisGlyPheLeuAspGlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
QY 2480 ATCCGAGCAGGAGAAACCTTACACAGCTCCAGATCTACCCCAAGACAGACAGATTTGCC 2539
Db 833 ValArgAlaGlyLYsProLYrAspLeuGlnIleTYrProGlnGlnArgHisSerIleArg 852
QY 2540 TGCCCGAGTCCGCGAGCAGCTATGAGTACAGTTGCTGCTGCTCATGAGAAATACCTC 2599
Db 853 ValProGlnSerGlyGlnHisTYrGlnLeuHisLeuLeuHisTYrLeuGlnGlnAsnLeu 872

RESULT 4
075273 PRELIMINARY: PRT: 508 AA.
AC 075273:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE R26984_1 (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Atlix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coeffield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Rommelspacher A., Arellano A., Montgomey M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.,
RT "Sequence analysis of a 2.5 kb region in 19p13.3."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005594; AAC33801.1;
DR MEROFS; S09.019;
DR InterPro; IPR002410; Pro_annotase.
DR InterPro; IPR000379; Ser_estrs_site.
DR PRINTS; PRO0793; PROAMNOTASE.
FT NON_TER 1 1

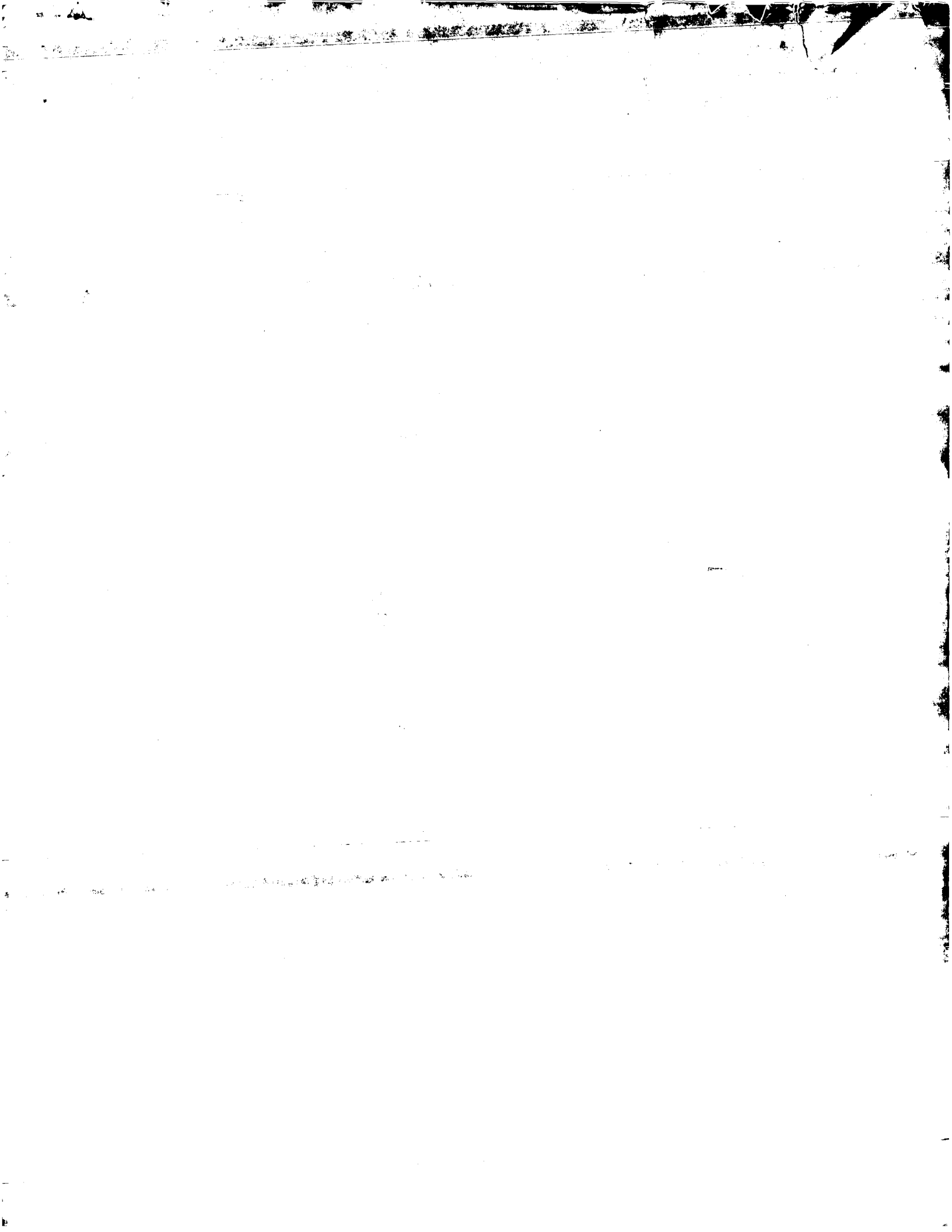
```

```

SQ. SEQUENCE 508 AA; 57750 MW; 2F4D6645BE2D2C89 CRC64;
Alignment Scores:
Pred. No.: 3.36e-163 Length: 508
Score: 2572.00 Matches: 493
Percent Similarity: 91.34% Conservative: 3
Best Local Similarity: 90.79% Mismatches: 3
Query Match: 53.64% Indels: 44
DB: 4 Gaps: 5

US-09-976-674-4 (1-2617) x 075273 (1-508)
QY 998 ATGCTGTGACCCAGAGAGAGAGAGAGTGGTGCACCCCTTACAGTGGCTTCCGAGAGTG 1057
Db 1 ILevalSerThrGlnGlnLYsGlnLeuValGlnProPheSerSerLeuProLYsVal 20
QY 1058 GAGTACATGCCAGGCGCGCGGTCACCCGCGGATGCAATACCGCTGGCGCAATGCTCTG 1117
Db 21 GlnTYrIleAlaArgAlaGly-----AlaTrpAlaMetPheLeu 33
QY 1118 GACCGGCGCCAGACAGTGGCTCAGCTGCTCTCTCCCGCGCTGTTCAATCCGAGAG 1177
Db 34 AspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProAlaLeuPheIleProSer 53
QY 1178 ACAGAGATGAGAGACAGCGGCTAGCGCTGCTGCGAGCTGTCCCGAGATGTCAGCGG 1237
Db 54 ThrGlnAsnGlnGlnGlnArgLeuAlaSerAlaArgAlaValProLYsAsnValGlnPro 73
QY 1238 TATGTGTGTACAGAGAGTCCACAGCTGTGATCAATGTTCATGACATTTCTATCC 1297
Db 74 TYrValValTYrGlnGlnValThrAsnValTrpIleAsnValHisAspIlePheTYrPro 93
QY 1298 TTCCCGCAATGAGAGAGAGAGAGAGAGTGTGCTTCCCGCGCAATGAAATGCAAGAGC 1357
Db 94 PheProGlnSerGlnGlnGlnAspLeuLeuLYsPheLeuArgAlaAsnGlyLYsTYr 113
QY 1358 GCGTCTGTGCATTTGTACAAAGTACCGCGCTTTAAATCCAGGCGTACGATTTGAGAT 1417
Db 114 GLYrPheLYsLeuTYrLYsValThrAlaValLeuLYsSerGlnIleTYrAspTrpSer 133
QY 1418 GAGCGCTTACGCGCGGAGAGATGAAATTAAGTCCCGCATTAAGAAAGATTGCTG 1477
Db 134 GlnProPheSerProGlnGlnGly-----GlnGlnSerLeu 145
QY 1478 ACCAGCGGTGAATGGAGAGTTTGGCGAGCGAGCTCAAGATCTGGTCAATGAGAG 1537
Db 146 ThrAsnAla-----IleTrpValAsnGlnGln 154
QY 1538 ACCAGCTGTGTACTTCCAGGCGACCAAGACAGCGCGCTGAGACACCACTCTACGTG 1597
Db 155 ThrLYsLeuValTYrPheGlnGlyThrLYsAspThrProLeuGlnHisLeuTYrVal 174
QY 1598 GTACAGTATGAGCGCGCGGAGATCTGACGCTCACAGCGCGCGCTCTCCATACG 1657
Db 175 ValSerTYrGlnAlaAlaGlyGlnIleValArgLeuThrThrProGlyPheSerHisSer 194
QY 1658 TGCCTCATGACGCAACAATCTGACATGTGCTGACCCACTACAGACGCTGACAGCGCG 1717
Db 195 CysSerMetSerGlnAspPheAspMetPheValSerHisTYrSerValSerThrPro 214
QY 1718 CCCTCGCTGACAGCTTAAAGTGAAGCGCGCGAGACGACGCGCGCTGACAGAGCGCC 1777
Db 215 ProGlyValHisValTYrLYsLeuSerGlyProAspAspProLeuHisLYsGlnPro 234
QY 1778 CGCTTCTGGGCTAGCATATGAGAGACAGCTGCGCGCGGATATATGTTCTCTCCAGAG 1837
Db 235 ArgPheTrpAlaSerMetGlnAlaAla-----LYs 245
QY 1838 ATCTTCATTTTCCACAGCGCGCTCGATGTGCGCTTACGAGATCTACAGAGCGCCAC 1897
Db 246 ILePheHisPheHisThrArgSerArgValArgLeuTYrGlyMetIleTYrLYsProHis 265
QY 1898 GCGTTGACAGCCAGGAGAGAGACACCCACGCTCTTTGTATATGAGAGCGCCAGAGTG 1957

```



Db 312 AspaspProLysSerAlaGlyValAlaLathrPheValIleGlnGlnLuphneAspArgPhe 331
 QY 770 ACTGGCTACTGTGGTGGCCACAGCCCTCGGAGAGCTTCACAGAGGCCCTCAAGAGCTG 829
 Db 332 ThrGlyTrpTrpPheCysProThrAlaSerTrpGlu-----GlnGlyLeuYsTrpIleu 349
 QY 830 CGAATCTGTATGAGGAAGTCCAGTCCAGAGTGGAGGTCAATTCAGTCCCTCTCT 889
 Db 350 ArgIleLeuYrGlnGlnValAlaSpGlnSerGlnValGlnValIleHisValProSerPro 369
 QY 890 GCGCTAGAGAAAGAGAGAGAGAGTCTGTATCGGTGAGTCCCGAGAGAGAGAGAGTCC 949
 Db 370 AlaLeuGlnGlnValGlnYsTrpAspSerTrpArgTrpArgTrpArgTrpArgTrpArgTrp 389
 QY 950 AAGATTCCTTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1009
 Db 390 LysIleAlaLeuLysLeuAlaGlnLuphGlnThrAspSerGlnGlnLysIleValSerThr 409
 QY 1010 CAGGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1069
 Db 410 GlnGlnLysGlnLuphValGlnProPheSerSerLeuPheProLysValGlnTrpIleAla 429
 QY 1070 AGGCGCCGG 1078
 Db 430 ArgAlaGly 432
 RESULT 8
 Q96NT8 PRELIMINARY; PRT; 312 AA.
 ID Q96NT8
 AC Q96NT8;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE CDNA FLJ30094 f1s, clone BNGH4100034, weakly similar to dipeptidyl
 DE peptidase IV (EC 3.4.14.5).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
 RA Ishii S., Yamamoto J., Isono Y., Kawai-Hiro Y., Saito K., Nishikawa T.,
 RA Kikuchi K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
 RA Takahashi H., Kanda K., Wagasuma M., Murakawa K., Kanehori K.,
 RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
 RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
 RT "MEDO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK054656; BAB70784.1;
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR00379; Ser esters_site.
 DR Pfam: PF00326; Peptidase_S9; 1
 SQ SEQUENCE 312 AA; 35518 MW; ABE940AFC5877717 CRC64;

QY 1784 TGGGCTACATGATGAG 1843
 Db 41 TrpAlaSerMetMetGlnAlaAlaSerCysProAspTrpValProProGlnIlePhe 60
 QY 1844 CATTCACAG 1903
 Db 61 HisPheHisThrArgSerSerPalaArgLeuYrGlyMetIleYrLysProHisAlaLeu 80
 QY 1904 CAGCAG 1963
 Db 81 GlnProGlyLysLysHisProThrValLeuPheValYrGlyGlyProGlnValGlnLeu 100
 QY 1964 GTGATTAACCTCTCAAG 2023
 Db 101 ValAsnAsnSerPheLysGlnYrIleYrLysLeuArgLeuAsnThrLeuAlaSerLeuGly 120
 QY 2024 TACGCGGTGGTGTGATGAG 2083
 Db 121 TyrAlaValAlaValAlaLeuSpGlyArgGlySerCysGlnArgGlyLeuArgPheGlnGly 140
 QY 2084 GCGCTGAAAACCAATGGGCGAG 2143
 Db 141 AlaLeuLysAsnGlnMetGlnValGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 160
 QY 2144 GTGCGGAGAGAGATGAGTTCATGACAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2203
 Db 161 ValAlaGlnLysTrpGlnLysPheLeuLeuSerAlaValAlaIleHisGlyTrpSerTrp 180
 QY 2204 GGGGCGTCTCTCTGCTCATGGGCGCTATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2263
 Db 181 GlyLysPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIle 200
 QY 2264 GCGGAGGCGCGCGTCCGCTGAG 2323
 Db 201 AlaCysAlaProValAlaThrValTrpMetAlaTrpAspTrpGlyTrpGlnArgTrpMet 220
 QY 2324 GACGTCCTGAG 2383
 Db 221 AspValProGlnAsnAsnGlnHisGlyTrpGlnAlaGlySerValAlaLeuHisValGln 240
 QY 2384 AAGCTCCCAATGAG 2443
 Db 241 LysLeuProAsnGlnLuphAsnArgLeuLeuIleHisGlyPheLeuAspGlnAsnVal 260
 QY 2444 CACTTTTCCACAAACTTCTCTGCTCCCAACTGATCCGAGAGAGAGAGAGAGAGAGAGAG 2503
 Db 261 HisPhePheHisThrAsnPheLeuValSerGlnLeuIleAlaGlnGlyLysProTyrGln 280
 QY 2504 CTCACATCTACCCCAAG 2563
 Db 281 LeuGlnIleYrProAsnGlnLuphArgHisSerIleArgCysProGlnSerGlyGlnHisTrp 300
 QY 2564 GAAGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2599
 Db 301 GlnValThrLeuLeuHisPheLeuGlnGlnLuphTrpLeu 312
 RESULT 9
 Q9VC20 PRELIMINARY; PRT; 1042 AA.
 ID Q9VC20
 AC Q9VC20;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE CG3744 protein.
 GN CG3744.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Eurygola; Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 11:53:37 ; Search time 34 Seconds
(without alignments)
3382.214 Million cell updates/sec

Title: US-09-976-674-3
Perfect score: .4646
Sequence: 1 MATTGTPRADRGDAATDPD.....CPESGEHYETLLHFLQEYL 863

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4646	100.0	863	23	ABG61592 Human DPPIV relate
2	4646	100.0	892	23	ABG61602 Human DPP-2 splic
3	4646	100.0	892	23	ABG61604 Human DPP-2 splic
4	4636	99.8	892	23	ABG61607 Human DPP-2 splic
5	4558.5	98.1	879	23	ABG61608 Human DPP-2 splic
6	4558.5	98.1	879	23	ABG61608 Human DPP-2 splic
7	4458	96.0	830	23	AAE24171 Human dipeptidyl p
8	4279	92.1	869	23	AAE24169 Alternative versio
9	4129	88.9	847	23	AAE23875 Murine dipeptidyl
10	4026	86.7	832	23	ABG61605 Human DPP-2 splic

11	4026	86.7	832	23	ABG61606 Human DPP-2 splic
12	3938.5	84.8	819	23	ABG61609 Human DPP-2 splic
13	3938.5	84.8	819	23	ABG61610 Human DPP-2 splic
14	3817.5	82.2	737	22	AAH38724 Human polypeptide
15	3539.5	76.2	683	22	AAH40510 Human polypeptide
16	3409.5	73.4	720	21	AAH41626 Human OREF ORF1390
17	2870	61.8	882	22	AAH47187 Human DPP8. Homo
18	2870	61.8	882	23	AAE24170 Human dipeptidyl p
19	2870	61.8	882	23	ABG61591 Human DPPIV relate
20	2870	61.8	882	23	AAU74749 Human protease PRT
21	2870	61.8	882	23	AAH78415 Human albumin fusi
22	2688	57.9	497	23	ABG64845 Human secreted pro
23	2688	57.9	497	23	AAU96169 Human albumin fusi
24	2684	57.8	497	23	ABG64844 Human albumin fusi
25	2684	57.8	497	23	AAU96192 Human secreted pro
26	2572	55.4	508	21	AAH42928 Human OREF ORF2692
27	2562	55.1	518	21	ABG61603 Human DPP-2 splic
28	2547	54.8	518	21	AAH90299 Human peptidase, H
29	2414	52.0	782	23	ABH97361 Novel human protei
30	2178	46.9	724	23	ABH97362 Novel human protei
31	2137.5	46.0	632	22	ABH33565 Human protein sequ
32	2109.5	45.4	460	22	AAH38740 Human polypeptide
33	2103.5	45.3	512	22	AAH40526 Human polypeptide
34	1998	43.0	690	23	ABG61594 Human DPP-1 splic
35	1997	43.0	658	23	ABG61600 Human DPP-1 splic
36	1997	43.0	661	23	ABG61596 Human DPP-1 splic
37	1816	39.1	613	23	ABG61601 Human DPP-1 splic
38	1628.5	35.1	1042	22	ABH60137 Drosophila melanog
39	1628.5	35.1	1102	22	ABH60209 Drosophila melanog
40	1583	34.1	580	23	AAE14337 Human DPP8 319hr-
41	1494.5	32.2	465	22	AAH47189 Human DPP8 524ph-
42	1220.5	26.3	360	22	AAH47188 Human dipeptidyl p
43	1093.5	23.5	310	22	ABH08994 Human albumin fusi
44	1093.5	23.5	310	23	ABH08994 Human albumin fusi
45	1047	22.5	349	23	ABG64842 Human albumin fusi

ALIGNMENTS

RESULT 1	ABG61592	standard; Protein: 863 AA.
ID	ABG61592	
XX	12-AUG-2002 (first entry)	
AC	ABG61592:	
XX	12-AUG-2002 (first entry)	
DT	Human DPPIV related serine protease DPP-2.	
XX		
XX	Human: serine protease; dipeptidyl peptidase IV-related protein; DPP:	
KW	DPPIV; Infection; human immunodeficiency virus; HIV-1; HIV-2; pain;	
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;	
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;	
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;	
KW	dyskinesia; reproductive disorder; inflammatory disorder;	
KW	metabolic disorder.	
XX		
OS	Homo sapiens.	
PN	WO20023134-A2.	
XX		
PD	18-APR-2002.	
XX		
PF	12-OCT-2001; 2001WO-US31874.	
XX		
PR	12-OCT-2000; 2000US-240117P.	
XX		
PA	(FERR) FERRING BV.	
XX		
PI	Ol S, Akinsanya KO, Riviere PJ, Junien J;	
XX	WPI; 2002-444178/A7.	

DR N-PSDB; ABR83323.
 XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 XX
 PS Claim 17, Fig 1: 113pp: English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPP-IV proteins.
 CC
 XX
 XX Sequence 863 AA:
 Query Match 100.0%; Score 4646; DB 23; Length 863;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MATGTGTADRGDAATDDPAARFOVQKHSMDGLRSIIHGRKXSGILVKNRAPHDFQVQ 60
 DB 1 MATGTGTADRGDAATDDPAARFOVQKHSMDGLRSIIHGRKXSGILVKNRAPHDFQVQ 60
 QY 61 KTDSSGSHSLYLLGMPYGRSLSLYSEIPKVRKALLLSMKOMLDFQATPHHGV 120
 DB 61 KTDSSGSHSLYLLGMPYGRSLSLYSEIPKVRKALLLSMKOMLDFQATPHHGV 120
 QY 121 YSRREELLRERKRLGVGITSYDFHSESGFLFOASNSLFFCRDGGKGFVWSPMKPLEI 180
 DB 121 YSRREELLRERKRLGVGITSYDFHSESGFLFOASNSLFFCRDGGKGFVWSPMKPLEI 180
 QY 121 YSRREELLRERKRLGVGITSYDFHSESGFLFOASNSLFFCRDGGKGFVWSPMKPLEI 180
 DB 121 YSRREELLRERKRLGVGITSYDFHSESGFLFOASNSLFFCRDGGKGFVWSPMKPLEI 180
 QY 181 KTCGSGPRMDKICPADPAFFSITNSDLWVANIETGEERLFTCHQGLSVLDDPKSAG 240
 DB 181 KTCGSGPRMDKICPADPAFFSITNSDLWVANIETGEERLFTCHQGLSVLDDPKSAG 240
 QY 241 VAFVIOEEPRFTGYWMCPTASWEGSGKTLRLIYEVDSEVEVLIHVPSPALEEKT 300
 DB 241 VAFVIOEEPRFTGYWMCPTASWEGSGKTLRLIYEVDSEVEVLIHVPSPALEEKT 300
 QY 241 VAFVIOEEPRFTGYWMCPTASWEGSGKTLRLIYEVDSEVEVLIHVPSPALEEKT 300
 DB 241 VAFVIOEEPRFTGYWMCPTASWEGSGKTLRLIYEVDSEVEVLIHVPSPALEEKT 300
 QY 301 DSYRPTGSKNPKIALKLAIEFQTDGKIVSTOEKELVQFSSLPKVEYIARAGWTRD 360
 DB 301 DSYRPTGSKNPKIALKLAIEFQTDGKIVSTOEKELVQFSSLPKVEYIARAGWTRD 360
 QY 301 DSYRPTGSKNPKIALKLAIEFQTDGKIVSTOEKELVQFSSLPKVEYIARAGWTRD 360
 DB 301 DSYRPTGSKNPKIALKLAIEFQTDGKIVSTOEKELVQFSSLPKVEYIARAGWTRD 360
 QY 361 GKIVAMAFLDPRQOMLOLVLLPALFTPTSTNEBQRLASARAVPRNOPYVVEEVTNW 420
 DB 361 GKIVAMAFLDPRQOMLOLVLLPALFTPTSTNEBQRLASARAVPRNOPYVVEEVTNW 420
 QY 421 INVNDITFPFQSGEDELCLFRANECKTGFCCHLYKVAVALKSGQYDMSSEFSGEDEEK 480
 DB 421 INVNDITFPFQSGEDELCLFRANECKTGFCCHLYKVAVALKSGQYDMSSEFSGEDEEK 480
 QY 481 CPIKEETALTGEMEVILARHSGKIWNNEETKLYVFOGTQKDPLEHNLVYVEAAGELVR 540
 DB 481 CPIKEETALTGEMEVILARHSGKIWNNEETKLYVFOGTQKDPLEHNLVYVEAAGELVR 540
 QY 541 LTPGFSHSCMSQNFDMFVSHYSSVSTPCVHVYKLSGDDDLHKQPRFWMASMEAS 600
 DB 541 LTPGFSHSCMSQNFDMFVSHYSSVSTPCVHVYKLSGDDDLHKQPRFWMASMEAS 600
 QY 601 CRRPYVPEETIFHFTSRDVRILYGMILYKPHALOPGKKHPTVLFVGGQVOLVNNSEFKIK 660
 DB 601 CRRPYVPEETIFHFTSRDVRILYGMILYKPHALOPGKKHPTVLFVGGQVOLVNNSEFKIK 660

QY 661 YLRNLTLASIGYAAVVVIDGRSCORGLRFEGALKNGQVEIEDQVGLQFPAEKYGTID 720
 DB 661 YLRNLTLASIGYAAVVVIDGRSCORGLRFEGALKNGQVEIEDQVGLQFPAEKYGTID 720
 QY 721 LSRVAIHGWSYGGFLSLMGLIHKPQVKVAIAGAPVYVMAAYDTGTYERYVDEENNQHG 780
 DB 721 LSRVAIHGWSYGGFLSLMGLIHKPQVKVAIAGAPVYVMAAYDTGTYERYVDEENNQHG 780
 QY 781 YEAGSVLHVKEKLPNEPRNRLILGFLDENVHFHTNPLVSQILRACKPYQLOIYPERH 840
 DB 781 YEAGSVLHVKEKLPNEPRNRLILGFLDENVHFHTNPLVSQILRACKPYQLOIYPERH 840
 QY 841 SIRCESEHNEYVTLHFLQEYL 863
 DB 841 SIRCESEHNEYVTLHFLQEYL 863
 RESULT 2
 ABG61602
 ID ABG61602 standard; Protein: 892 AA.
 XX
 AC ABG61602;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPP-2 splice variant #1.
 XX
 KW Human: serine protease; dipeptidyl peptidase IV-related protein; DPP-IV; DPP-IV; Infection: human immunodeficiency virus; HIV-1; HIV-2; pain; diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke; heart failure; hypertension; urinary retention; osteoporosis; cancer; ulcer; allergy; cancer; psychotic disorder; neurological disorder; dyskinesia; reproductive disorder; inflammatory disorder; metabolic disorder.
 KW
 OS Homo sapiens.
 XX
 PN W0200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junten J;
 XX
 DR WPI: 2002-444178/47.
 DR
 DR N-PSDB; ABR83333.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 XX
 PS Disclosure; Page 76-78; 113pp: English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and

CC	metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX	
XX	
Sequence	892 AA;
Query Match	100.0%; Score 4646; DB 23; Length 892;
Best Local Similarity	100.0%; Pred. No. 0;
Matches	863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MATGTPTADRDGAATDDPAARFOVOKHSMGDLSTIHSKRKYSGLLVNKAHPHFOFVO 60
Db	30 MATGTPTADRDGAATDDPAARFOVOKHSMGDLSTIHSKRKYSGLLVNKAHPHFOFVO 89
Qy	61 KTDESGPHSHRLYLGMYPGSRNSLXSEIPKVKRKALLLSMKOMLDHFOATPHHGV 120
Db	90 KTDESGPHSHRLYLGMYPGSRNSLXSEIPKVKRKALLLSMKOMLDHFOATPHHGV 149
Qy	121 YSRBEELRERKRLGVFGITSDYDFHSESGFLFOASNSLFHCRDGGKNGFVSPKPLEI 180
Db	150 YSRBEELRERKRLGVFGITSDYDFHSESGFLFOASNSLFHCRDGGKNGFVSPKPLEI 209
Qy	181 KTQCSGPRMDPKICPADPAFFSFINNSDLVANITGEERRLTFCGGLSNVLDPPKSNAG 240
Db	210 KTQCSGPRMDPKICPADPAFFSFINNSDLVANITGEERRLTFCGGLSNVLDPPKSNAG 269
Qy	241 VATEVIOREDFRTGYWMCPTASWEGSEGLKTLRLILEVDESEVEVLIHVSPPALREKRT 300
Db	270 VATEVIOREDFRTGYWMCPTASWEGSEGLKTLRLILEVDESEVEVLIHVSPPALREKRT 329
Qy	301 DSYRPTGTGSKNPRIALKAELAEFQDSDGKIVSTOEKELVQPFSSLPKVEYIARAGWTRD 360
Db	330 DSYRPTGTGSKNPRIALKAELAEFQDSDGKIVSTOEKELVQPFSSLPKVEYIARAGWTRD 389
Qy	361 GKVAMAMLDPRQOMLOLVLLPPLAFITSTENEBQRLASARAVRPNNOPIVVEVTVW 420
Db	390 GKVAMAMLDPRQOMLOLVLLPPLAFITSTENEBQRLASARAVRPNNOPIVVEVTVW 449
Qy	421 INVADIEFPQSEGEDELCEFLRANECKTGFCCHLYKVTAVLSQGYDMSPEPSEDEFEK 480
Db	450 INVADIEFPQSEGEDELCEFLRANECKTGFCCHLYKVTAVLSQGYDMSPEPSEDEFEK 509
Qy	481 CPIKEEIALTSGEWEVLARHSGSKIWNNEETKLVYFQGTDPRLHNHLYVSYEAGETIVR 540
Db	510 CPIKEEIALTSGEWEVLARHSGSKIWNNEETKLVYFQGTDPRLHNHLYVSYEAGETIVR 569
Qy	541 LTTGFGFSSCMSQNFEDFVSHSVSTPPCVHYKLSGPDDEPLHKQPRFASMMEAAS 600
Db	570 LTTGFGFSSCMSQNFEDFVSHSVSTPPCVHYKLSGPDDEPLHKQPRFASMMEAAS 629
Qy	601 CPPDVPEPEIFHFTRSQVRLYGMITYKPHALOPGKKHPTVLFVYGGPOVOLVNNSEFKGIK 660
Db	630 CPPDVPEPEIFHFTRSQVRLYGMITYKPHALOPGKKHPTVLFVYGGPOVOLVNNSEFKGIK 689
Qy	661 YLRNLTLASLIGYAVVVVIDRGSCQGLRFEGLAKNMQOVETEDVESLQFVAEKYGEFLD 720
Db	690 YLRNLTLASLIGYAVVVVIDRGSCQGLRFEGLAKNMQOVETEDVESLQFVAEKYGEFLD 749
Qy	721 LSRVAIHMSWGGLSLMGLIHKROYEVKATAGAPYVMAVDTGYTERKYDVPENNCHG 780
Db	750 LSRVAIHMSWGGLSLMGLIHKROYEVKATAGAPYVMAVDTGYTERKYDVPENNCHG 809
Qy	781 YEAGSVALHVEKLEPNERRLILHGFLEENYHFFHTNLFVLSOLIRAGRPYOLQIYPNERH 840
Db	810 YEAGSVALHVEKLEPNERRLILHGFLEENYHFFHTNLFVLSOLIRAGRPYOLQIYPNERH 869
Qy	841 SIRPESGEHYEVTLLHFLQEYL 863
Db	870 SIRPESGEHYEVTLLHFLQEYL 892
RESULT 3	
ABG61604	
ID	ABG61604 standard; Protein: 892 AA.
XX	

AC	ABG61604;
XX	
XX	
DT	12-AUG-2002 (first entry)
DE	
XX	
XX	
KW	Human DPRP-2 splice variant #3.
KW	Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW	DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW	diabetes; reproductive disorder; inflammatory disorder;
KW	metabolic disorder.
OS	
XX	Homo sapiens.
XX	
XX	WO200231134-A2.
XX	
XX	18-APR-2002.
XX	
XX	12-OCT-2001; 2001WO-US31874.
XX	
XX	12-OCT-2000; 2000US-240117P.
XX	
XX	(FERR) FERRING BV.
XX	
XX	Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX	
XX	WPI; 2002-444178/47.
XX	
XX	N-PSDB; ABK83335.
XX	
XX	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
XX	the proteins, useful for treating e.g. fungal, bacterial, protozoan and
XX	viral infections, cancers, allergies, neurological disorders, or pain
XX	
XX	
XX	Disclosure: Page 81-84; 113pp; English.
XX	
XX	
XX	The present invention relates to the isolation of novel human serine
XX	proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
XX	proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
XX	and nucleic acids encoding them are useful for treating infections
XX	such as fungal, bacterial, protozoan and viral infections, particularly
XX	infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
XX	pain, diabetes, precocious puberty, infertility, obesity, anorexia,
XX	bulimia, Parkinson's disease, acute heart failure, hypertension,
XX	hypertension, urinary retention, osteoporosis, angina pectoris,
XX	stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
XX	psychotic and neurological disorders (e.g. anxiety, dementia, or
XX	schizophrenia), and dyskinesias. These may also be used in discovering
XX	therapeutic agents for the treatment of reproductive, inflammatory and
XX	metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX	
XX	
Sequence	892 AA;
Qy	
Query Match	100.0%; Score 4646; DB 23; Length 892;
Best Local Similarity	100.0%; Pred. No. 0;
Matches	863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MATGTPTADRDGAATDDPAARFOVOKHSMGDLSTIHSKRKYSGLLVNKAHPHFOFVO 60
Db	30 MATGTPTADRDGAATDDPAARFOVOKHSMGDLSTIHSKRKYSGLLVNKAHPHFOFVO 89
Qy	61 KTDESGPHSHRLYLGMYPGSRNSLXSEIPKVKRKALLLSMKOMLDHFOATPHHGV 120
Db	90 KTDESGPHSHRLYLGMYPGSRNSLXSEIPKVKRKALLLSMKOMLDHFOATPHHGV 149
Qy	121 YSRBEELRERKRLGVFGITSDYDFHSESGFLFOASNSLFHCRDGGKNGFVSPKPLEI 180
Db	150 YSRBEELRERKRLGVFGITSDYDFHSESGFLFOASNSLFHCRDGGKNGFVSPKPLEI 209
Qy	181 KTQCSGPRMDPKICPADPAFFSFINNSDLVANITGEERRLTFCGGLSNVLDPPKSNAG 240
XX	

```
Db 210 KTQCSGPRMDPKICPADPAFFSFINNNDLWVANIETGEERLTFCHQGLSNVLDDPKSAG 269
Qy 241 VATEVIOEEDFRFTGYWMCPTASMEGSEGLKTLRLIYEEDVESEVEVYIHVSPALAEKRT 300
Db 270 VATEVIOEEDFRFTGYWMCPTASMEGSEGLKTLRLIYEEDVESEVEVYIHVSPALAEKRT 329
Qy 301 DSYRYPRTGSKNPKIALKLAIEFQTDGSKIVSTQEKELVOPFSSLPKVEYIARAGWTRD 360
Db 330 DSYRYPRTGSKNPKIALKLAIEFQTDGSKIVSTQEKELVOPFSSLPKVEYIARAGWTRD 389
Qy 361 GKYAMAMFLDRPQOMLOLVLLPPLALFIPSTENEBORLASARAVPRNVPYVVEEVNTNW 420
Db 390 GKYAMAMFLDRPQOMLOLVLLPPLALFIPSTENEBORLASARAVPRNVPYVVEEVNTNW 449
Qy 421 INVHDIFFPPOSGEDELCLFLRANECKTGCFCHLYKTVAVLKSOGYDMSSEPFSGEDEEFK 480
Db 450 INVHDIFFPPOSGEDELCLFLRANECKTGCFCHLYKTVAVLKSOGYDMSSEPFSGEDEEFK 509
Qy 481 CPIKEEIALTSGEWEVLARHSGKIVNNEETKLVYFOGTRDPLRHHLLYVSYEAAGETIVR 540
Db 510 CPIKEEIALTSGEWEVLARHSGKIVNNEETKLVYFOGTRDPLRHHLLYVSYEAAGETIVR 569
Qy 541 LITPGFSHSCMSQNFDMFVSHYSSVSTPCVHYVYKLSGPDDDLHROPREFWASMEAS 600
Db 570 LITPGFSHSCMSQNFDMFVSHYSSVSTPCVHYVYKLSGPDDDLHROPREFWASMEAS 629
Qy 601 CPPDYVPPEIFHFHTRSDVRLYGMITYKPHALOPGKHPVLFVYGGPOVOLVNNSEFKIK 660
Db 630 CPPDYVPPEIFHFHTRSDVRLYGMITYKPHALOPGKHPVLFVYGGPOVOLVNNSEFKIK 689
Qy 661 YLRILNTLASLGYAVVYIDGRSCQGRLEFEGALKNQMGVEIEDVQEGLOFAEKYGFID 720
Db 690 YLRILNTLASLGYAVVYIDGRSCQGRLEFEGALKNQMGVEIEDVQEGLOFAEKYGFID 749
Qy 721 LSRVAIHMSVSGFLSLMGLLHKPQVEKATAGAPVYVMAVYDGYTERYMDVPENNHG 780
Db 750 LSRVAIHMSVSGFLSLMGLLHKPQVEKATAGAPVYVMAVYDGYTERYMDVPENNHG 809
Qy 781 YEAGSVLHVKEKLPNEPRRLILHGFIDENVHFHTNPLVLSQILRAGKPYOLOIYPRNRH 840
Db 810 YEAGSVLHVKEKLPNEPRRLILHGFIDENVHFHTNPLVLSQILRAGKPYOLOIYPRNRH 869
Qy 841 SIRCESEGEHEVTLHFLQEYL 863
Db 870 SIRCESEGEHEVTLHFLQEYL 892

RESULT 4
AAE24168
ID AAE24168 standard; Protein: 969 AA.
XX
AC AAE24168;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 9 (DPP9) protein.
XX
KM Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KM autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KM graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KM antiviral; enzyme.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 374 /note="Encoded by GAA"
XX
XX WO200234900-A1.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-AU01388.
```

```
XX
PR 27-OCT-2000; 2000AU-0001078.
XX
PA (UNSY ) UNIV SYDNEY.
XX
PI Abbott CA, Gorrell MD;
XX
DR WPI; 2002-454646/48.
DR N-PSDB; AAD38954.
XX
FT New dipeptidyl peptidase (DPP) peptidases, useful for screening
FT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
PS Claim 1; Fig 4; 91pp; English.
XX
CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptidases are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP9 protein.
XX
SQ Sequence 969 AA;
XX
Query Match 99.8%; Score 4636; DB 23; Length 969;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MATTGTPADRGDAATDDPARFOYOKHSDGLRSIIHSGRSKYSGLIVNKAPHDQFQV 60
Db 107 MATTGTPADRGDAATDDPARFOYOKHSDGLRSIIHSGRSKYSGLIVNKAPHDQFQV 166
Qy 61 KTDSGSPSHSLYLYLGMFYSGRENSLSYSEIPKVKYRKALLLSMKQMLDFOATPHHGV 120
Db 167 KTDSGSPSHSLYLYLGMFYSGRENSLSYSEIPKVKYRKALLLSMKQMLDFOATPHHGV 226
Qy 121 YSRREELLREKRLGVEGITSYDHSSEGLPLFOASNSLFCRDRGKNGFVSPMKPLEI 180
Db 227 YSRREELLREKRLGVEGITSYDHSSEGLPLFOASNSLFCRDRGKNGFVSPMKPLEI 286
Qy 181 KTQCSGPRMDPKICPADPAFFSFINNNDLWVANIETGEERLTFCHQGLSNVLDDPKSAG 240
Db 287 KTQCSGPRMDPKICPADPAFFSFINNNDLWVANIETGEERLTFCHQGLSNVLDDPKSAG 346
Qy 241 VATEVIOEEDFRFTGYWMCPTASMEGSEGLKTLRLIYEEDVESEVEVYIHVSPALAEKRT 300
Db 347 VATEVIOEEDFRFTGYWMCPTASMEGSEGLKTLRLIYEEDVESEVEVYIHVSPALAEKRT 406
Qy 301 DSYRYPRTGSKNPKIALKLAIEFQTDGSKIVSTQEKELVOPFSSLPKVEYIARAGWTRD 360
Db 407 DSYRYPRTGSKNPKIALKLAIEFQTDGSKIVSTQEKELVOPFSSLPKVEYIARAGWTRD 466
Qy 361 GKYAMAMFLDRPQOMLOLVLLPPLALFIPSTENEBORLASARAVPRNVPYVVEEVNTNW 420
Db 467 GKYAMAMFLDRPQOMLOLVLLPPLALFIPSTENEBORLASARAVPRNVPYVVEEVNTNW 526
Qy 421 INVHDIFFPPOSGEDELCLFLRANECKTGCFCHLYKTVAVLKSOGYDMSSEPFSGEDEEFK 480
Db 527 INVHDIFFPPOSGEDELCLFLRANECKTGCFCHLYKTVAVLKSOGYDMSSEPFSGEDEEFK 586
Qy 481 CPIKEEIALTSGEWEVLARHSGKIVNNEETKLVYFOGTRDPLRHHLLYVSYEAAGETIVR 540
Db 587 CPIKEEIALTSGEWEVLARHSGKIVNNEETKLVYFOGTRDPLRHHLLYVSYEAAGETIVR 646
Qy 541 LITPGFSHSCMSQNFDMFVSHYSSVSTPCVHYVYKLSGPDDDLHROPREFWASMEAS 600
Db 647 LITPGFSHSCMSQNFDMFVSHYSSVSTPCVHYVYKLSGPDDDLHROPREFWASMEAS 706
Qy 601 CPPDYVPPEIFHFHTRSDVRLYGMITYKPHALOPGKHPVLFVYGGPOVOLVNNSEFKIK 660
Db 707 CPPDYVPPEIFHFHTRSDVRLYGMITYKPHALOPGKHPVLFVYGGPOVOLVNNSEFKIK 766
```

Qy	661	YLRLNTLASLGVAVVYIDGRSGCQGLRPEGALKKMQMQVEIEDVEGLQIAEAYKGFID	720
Db	767	YLRLNTLASLGVAVVYIDGRSGCQGLRPEGALKKMQMQVEIEDVEGLQIAEAYKGFID	826
Qy	721	LSRVAIHGMSYGGFSLMGLIHKPOVFYVAVLAGAVTVMVAVDGYTEHYMDVPPNNHG	780
Db	827	LSRVAIHGMSYGGFSLMGLIHKPOVFYVAVLAGAVTVMVAVDGYTEHYMDVPPNNHG	886
Qy	781	YEAGSVALHVEKLPNEPNRLLIHGFEDENVAFFFTNLFVSQLIAGKRPYQLOIIPNESH	840
Db	887	YEAGSVALHVEKLPNEPNRLLIHGFEDENVAFFFTNLFVSQLIAGKRPYQLOIIPNESH	946
Qy	841	SIRCEPGEHVEYTLIHLFQETL	863
Db	947	SIRCEPGEHVEYTLIHLFQETL	969
RESULT 5			
ID	ABG61607	standard; Protein; 879 AA.	
XX	ABG61607;		
XX	12-AUG-2002	(first entry)	
DE	Human DPRP-2 splice variant #6.		
XX	Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;		
KW	DPRIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;		
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;		
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;		
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;		
KW	dyskinesia; reproductive disorder; inflammatory disorder;		
KW	metabolic disorder.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200231134-A2.		
PD	18-APR-2002.		
XX			
PE	12-OCT-2001; 2001WO-US31874.		
PR	12-OCT-2000; 2000US-240117P.		
XX			
PA	(FERR) FERRING BV.		
PI	Qi S, Akinsanya KO, Riviere PJ, Junien J;		
XX	WPI: 2002-444178/47.		
DR	N-PSDB; ABR83338.		
XX			
PT	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding		
PT	the proteins, useful for treating e.g. fungal, bacterial, protozoan and		
PT	viral infections, cancers, allergies, neurological disorders, or pain		
PT	-		
XX			
PS	Disclosure: Page 91-93; 113pp; English.		
XX			
CC	The present invention relates to the isolation of novel human serine		
CC	proteases referred to as dipeptidyl peptidase IV (DPRIV)-related		
CC	proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)		
CC	and nucleic acids encoding them are useful for treating infections		
CC	such as fungal, bacterial, protozoan and viral infections, particularly		
CC	infections caused by human immunodeficiency virus (HIV-1 or HIV-2),		
CC	pain, diabetes, precocious puberty, infertility, obesity, anorexia,		
CC	bulimia, Parkinson's disease, acute heart failure, hypotension,		
CC	hypertension, urinary retention, osteoporosis, and/or pectoris,		
CC	stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,		
CC	psychotic and neurological disorders (e.g. anxiety, dementia, or		
CC	schizophrenia), and dyskinesias. These may also be used in discovering		
CC	therapeutic agents for the treatment of reproductive, inflammatory and		

CC	metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX	Sequence 879 AA;
XX	Query Match 98.1%; Score 4558.5; DB 23; Length 879; Best Local Similarity 98.5%; Pred. No. 0; Matches 850; Conservative 0; Mismatches 0; Indels 13; Gaps 1;
QY	1 MATGTPTADRDAAATDPPARFQVOKHSMDSGLSIIHGSKYSGLLIVNKAHPDFQFVQ 60
Db	30 MATGTPTADRDAAATDPPARFQVOKHSMDSGLSIIHGSKYSGLLIVNKAHPDFQVQ 89
QY	61 KTDESGPSSHRLYLIGMPYSGRENSLLSEIRPKVKREALLLSWKOMLDHFQATPHNGV 120
Db	90 KTDESGPSSHRLYLIGMPYSGRENSLLSEIRPKVKREALLLSWKOMLDHFQATPHNGV 149
QY	121 YSRREELLRRRRRLGVFGITSYDFHSESGLLFQASNSLFLFCROGKKGFWSPKRLPI 180
Db	150 YSRREELLRRRRRLGVFGITSYDFHSESGLLFQASNSLFLFCROGKKGFWSPKRLPI 209
QY	181 KTQCGSPHMDPKICPADPAFFSFINNSDLVANIETGGERLLTFCHQSLSNVLDPKSAQ 240
Db	210 KTQCGSPHMDPKICPADPAFFSFINNSDLVANIETGGERLLTFCHQSLSNVLDPKSAQ 269
QY	241 VAFVYIOEBDFRTGYWMCPTASWEGSGELKTLRIILEVEDESEVEYLIVHPSPALEERT 300
Db	270 VAFVYIOEBDFRTGYWMCPTASWEGSGELKTLRIILEVEDESEVEYLIVHPSPALEERT 329
QY	301 DSYRTPRPGSKMPKIALKLAERQTSQGLYVTOEKELVORPSSLFPRVEYIARAGWTRD 360
Db	330 DSYRTPRPGSKMPKIALKLAERQTSQGLYVTOEKELVORPSSLFPRVEYIARAGWTRD 389
QY	361 GKYAAMFLDRQOOLQVLRLPALFIRSTENEEORLIASARVPRNVORPYVVEEVTVM 420
Db	390 GKYAAMFLDRQOOLQVLRLPALFIRSTENEEORLIASARVPRNVORPYVVEEVTVM 449
QY	421 INVHDIFFYPPQSEGEDELCLFLANDECKTGFCHLKYTAVALKSQGYDMSPEPSDEDFK 480
Db	450 INVHDIFFYPPQSEGEDELCLFLANDECKTGFCHLKYTAVALKSQGYDMSPEPSDEDFK 509
QY	481 CPIKEEIALTSGEWEVLARHGSKIWNEETKLVYFGTKDPRLENNHLLVVSYEAGGEIVR 540
Db	510 CPIKEEIALTSGEWEVLARHGSKIWNEETKLVYFGTKDPRLENNHLLVVSYEAGGEIVR 556
QY	541 LITTPGSHSCSOMNDMVSHTSSVSPRCVHYVTKLSGDDDDRLHKOPRFASMEEAS 600
Db	557 LITTPGSHSCSOMNDMVSHTSSVSPRCVHYVTKLSGDDDDRLHKOPRFASMEEAS 616
QY	601 CPPDVVPEPIFFPHFRSDVRLYGLMYKPHALDPGKKHPVULFVYSGPOVLVNNSPFKIK 660
Db	617 CPPDVVPEPIFFPHFRSDVRLYGLMYKPHALDPGKKHPVULFVYSGPOVLVNNSPFKIK 676
QY	661 YLRLLTSLASLGAUVVYIDGRGSCQGLRREGALKNOMQGELEIDQVEGLQFAEKYGEID 720
Db	677 YLRLLTSLASLGAUVVYIDGRGSCQGLRREGALKNOMQGELEIDQVEGLQFAEKYGEID 736
QY	721 LSRVALIHGHSYGGFSLIMGLIHKPOVFKALAGAPYVYMAADDTGTYTERKYMVPENNOHG 780
Db	737 LSRVALIHGHSYGGFSLIMGLIHKPOVFKALAGAPYVYMAADDTGTYTERKYMVPENNOHG 796
QY	781 YEAGSVALLVEKLLPNEPNLLILHGFIDENVAHFPTNLFVLSQLIRAGRPYQLQIYPNBRH 840
Db	797 YEAGSVALLVEKLLPNEPNLLILHGFIDENVAHFPTNLFVLSQLIRAGRPYQLQIYPNBRH 856
QY	841 SIRCPESGEHYEVTLIHLQEYL 863
Db	857 SIRCPESGEHYEVTLIHLQEYL 879
XX	RESULT 6 ABG61608 ID ABG61608 standard; Protein: 879 AA.

AC ABG61608;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-2 splice variant #7.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPiV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinestia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
PI Q1 S, AKinsanya KO, Riviere PJ, Junien J;
XX
DR WPI: 2002-444178/47.
XX
DR N-PSDB; ABR83339.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
XX
PS Disclosure; Page 95-96; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPiV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 879 AA;
Query Match 98.1%; Score 4558.5; DB 23; Length 879;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 850; Conservative 0; Mismatches 0; Indels 13; Gaps 1;
QY 1 MATGTCTPADDGDAATDDPAARQVQKHSMDGLRSITHSRKTSGLIVNKAHPDFQVQ 60
DB 30 MATGTCTPADDGDAATDDPAARQVQKHSMDGLRSITHSRKTSGLIVNKAHPDFQVQ 89
QY 61 KTDSEGPSHRLTYLGMFGYSGRENSLYSEIPKVKRKALLLSMKOMLDFQATPHHGV 120
DB 90 KTDSEGPSHRLTYLGMFGYSGRENSLYSEIPKVKRKALLLSMKOMLDFQATPHHGV 149
QY 121 YSREELLREKRLGVGITSYDFHSESGFLFQASNSLFCRQDGGKGFVSPMKPLEI 180
DB 150 YSREELLREKRLGVGITSYDFHSESGFLFQASNSLFCRQDGGKGFVSPMKPLEI 209
QY 181 KTCGSGPRMDKICPADPAFFSTINSDLVANITGGERLFTCHQGLSVNLDPSKAG 240
|||||

DB 210 KTCGSGPRMDKICPADPAFFSTINSDLVANITGGERLFTCHQGLSVNLDPSKAG 269
QY 241 VATFVIOEFPDRFTGYMWCPTASMEGSEGLKTLRLTYEEDSEVEYVHVSPALEEKT 300
DB 270 VATFVIOEFPDRFTGYMWCPTASMEGSEGLKTLRLTYEEDSEVEYVHVSPALEEKT 329
QY 301 DSYRYPRTSGKNPKIALKLAEFQDSDQKIVSTQEKELVQPFSSLPKVEYIARAGWTRD 360
DB 330 DSYRYPRTSGKNPKIALKLAEFQDSDQKIVSTQEKELVQPFSSLPKVEYIARAGWTRD 389
QY 361 GKYAMAMFLDRPQOMLQVLLPPLFTSTENEGORLASAAVARNQPYVVEYVW 420
DB 390 GKYAMAMFLDRPQOMLQVLLPPLFTSTENEGORLASAAVARNQPYVVEYVW 449
QY 421 INVHDIFFPQSGEDELCTFLRANECKTGCHLYKYTAVALKSOGYDMSEPFSGEDEFK 480
DB 450 INVHDIFFPQSGEDELCTFLRANECKTGCHLYKYTAVALKSOGYDMSEPFSGEDEFK 509
QY 481 CPIKEEIALTSGEWEVLARHGSKIWNNEETKLVYFOGTRDPLEHNLVVSYEAGELIVR 540
DB 510 CPIKEEIALTSGEWEVLARHGSK-----GTRDPLEHNLVVSYEAGELIVR 556
QY 541 LTTGCFSHSCMSQNFQDFVSHYSYSTPCVHYVYKLSGPDDDLHKQPRFMAEMEAAS 600
DB 557 LTTGCFSHSCMSQNFQDFVSHYSYSTPCVHYVYKLSGPDDDLHKQPRFMAEMEAAS 616
QY 601 CPPDYVPEIHFHTRSDVRLYGMIKRPHALQPKKHTVLFYVGGPOVOLVNNSPFKIG 660
DB 617 CPPDYVPEIHFHTRSDVRLYGMIKRPHALQPKKHTVLFYVGGPOVOLVNNSPFKIG 676
QY 661 YLRNLTASLIGYAAVVVIDGRSCQGLRFEALKNOMQVEIEDQVEGLQFAEKYGFID 720
DB 677 YLRNLTASLIGYAAVVVIDGRSCQGLRFEALKNOMQVEIEDQVEGLQFAEKYGFID 736
QY 721 LSRVAIHGMSYSGFLSLMGLLHKPOVFYVATAGAPVYVMAVDPGYTRYKYDVEENQHG 780
DB 737 LSRVAIHGMSYSGFLSLMGLLHKPOVFYVATAGAPVYVMAVDPGYTRYKYDVEENQHG 796
QY 781 YEAGSVLHVKEKLPNEPRRLILHGFIDENYHFFTNFLVSQLIRAGKPYOLOIYPNERH 840
DB 797 YEAGSVLHVKEKLPNEPRRLILHGFIDENYHFFTNFLVSQLIRAGKPYOLOIYPNERH 856
QY 841 SIRCESEHYEVTLLHFLQEYL 863
DB 857 SIRCESEHYEVTLLHFLQEYL 879
RESULT 7
AAE24171
ID AAE24171 standard; Protein: 830 AA.
XX
AC AAE24171;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 4 (DPP4)-like 2 protein.
XX
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme; DPP-4 like 2 protein.
XX
OS Homo sapiens.
XX
FT Key Location/Qualifiers
FT MISC-difference 235 /note- "Encoded by GAG"
PD WO200234900-A1.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-AU01388.

• , Fri Dec 13 12:15:05 2002

```

XX 27-OCT-2000: 2000AU-0001078.
PR
XX
XX (UNSY ) UNIV SYDNEY.
PA
XX
XX Abbott CA, Gorrell MD;
XX
XX WPI: 2002-454646/48.
DR
XX N-PSDB: AAD38957.
XX
PT New dipeptidyl peptidase (DPP) peptidases, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
XX
PS Disclosure; Page 82-86; 91pp; English.
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX polynucleotides encoding such proteins. The DPP peptidases are useful for
XX screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV (human immuno deficiency virus) infection. The present
XX sequence is human DPP4-like 2 protein.
XX
SQ Sequence 830 AA;
Query Match 96.0%; Score 4458; DB 23; Length 830;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 828; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 34 LRSTIHSGSRKYSGLIVNKAPHDFQVOKTDESGPHSHRLYLYGMFYSGRENSLYSEIIPK 93
DB 1 LRSTIHSGSRKYSGLIVNKAPHDFQVOKTDESGPHSHRLYLYGMFYSGRENSLYSEIIPK 60
QY 94 KYRKEALLLSWKOMLDHFOATPHHGVYSREELLREKRRLGVFGITSYDFHSSGGLFLF 153
DB 61 KYRKEALLLSWKOMLDHFOATPHHGVYSREELLREKRRLGVFGITSYDFHSSGGLFLF 120
QY 154 QASNSLPHCRDGGKNGFMVSPMKLEIKTQCSGPRMDPKTCPPAPAFSFINNSDLWYAN 213
DB 121 QASNSLPHCRDGGKNGFMVSPMKLEIKTQCSGPRMDPKTCPPAPAFSFINNSDLWYAN 180
QY 214 IETGEERRLTFCHQGLSNVLDPPKSAGVAFVIOEEFDRTGYWMCPTASWEGSEGLTKL 273
DB 181 IETGEERRLTFCHQGLSNVLDPPKSAGVAFVIOEEFDRTGYWMCPTASWEGSEGLTKL 240
QY 274 RILVEEVDESEVEYIHWSPALERTDSTRYPRTGSKNPKIALKLAFOQDSOGKIYST 333
DB 241 RILVEEVDESEVEYIHWSPALERTDSTRYPRTGSKNPKIALKLAFOQDSOGKIYST 300
QY 334 QEKELVOPFSSLFPKVEYIAAGWTRDQKIVAMFLDRPQOMLDLVLLPALFTPSTENE 393
DB 301 QEKELVOPFSSLFPKVEYIAAGWTRDQKIVAMFLDRPQOMLDLVLLPALFTPSTENE 360
QY 394 EORLASAAVPRNQPYVYVEEVNWNVINVDIFYPPQSGEDELCLFLANECKTGCH 453
DB 361 EORLASAAVPRNQPYVYVEEVNWNVINVDIFYPPQSGEDELCLFLANECKTGCH 420
QY 454 LYKTAVALKSGQIDWSEPFSGEDEFKCPKIEEIALTSGEWEVLARHGSKIWNVEETKLV 513
DB 421 LYKTAVALKSGQIDWSEPFSGEDEFKCPKIEEIALTSGEWEVLARHGSKIWNVEETKLV 480
QY 514 YFOGTKTPTLEHNHLYVYSYEAGEIVRLTPPGFSHSCSMQNFPMFVSHSVSTPQVH 573
DB 481 YFOGTKTPTLEHNHLYVYSYEAGEIVRLTPPGFSHSCSMQNFPMFVSHSVSTPQVH 540
QY 574 VYKLSGDDPDLHKOPRFWASMMMEASCPDYPPELTFHFTTRSDVRLYGIYRPHALOP 633
DB 541 VYKLSGDDPDLHKOPRFWASMMMEASCPDYPPELTFHFTTRSDVRLYGIYRPHALOP 600
QY 634 GKRRPVLVYVYGGPOVOLVNNSEFKGITYRLNLNTLASLGAVAVVIDGRSGCORGLRFEGL 693
DB 601 GKRRPVLVYVYGGPOVOLVNNSEFKGITYRLNLNTLASLGAVAVVIDGRSGCORGLRFEGL 660

```

```

QY 694 KNMGQVEIEDQVEGLQFAEKYGFIDLSRYAIGHMSYGGFLSLMGLIHKPOVEKVALAT 753
DB 661 KNMGQVEIEDQVEGLQFAEKYGFIDLSRYAIGHMSYGGFLSLMGLIHKPOVEKVALAT 720
QY 754 APVTVMAYDYGYTERYNDVPENNQHGAEAGSVALHVEKLNEPNRLLILHGFDENYHF 813
DB 721 APVTVMAYDYGYTERYNDVPENNQHGAEAGSVALHVEKLNEPNRLLILHGFDENYHF 780
QY 814 FHTNPLVSQILRAGKPYOLQIYPNERHSIRCPESGEHEVTLHFLQEYL 863
DB 781 FHTNPLVSQILRAGKPYOLQIYPNERHSIRCPESGEHEVTLHFLQEYL 830

RESULT 8
AAE24169
ID AAE24169 standard; Protein; 869 AA.
XX
AC AAE24169;
XX
DT 23-SEP-2002 (first entry)
XX
DE Alternative version of murine dipeptidyl peptidase 9 (DPP9) protein.
XX
XX Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
XX human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
XX type II diabetes; antidiabetic; antiinflammatory; Immunosuppressive;
XX antiviral; enzyme.
XX
XX Mus sp.
XX
XX WO200234900-A1.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-AU01388.
XX
XX 27-OCT-2000; 2000AU-0001078.
XX
XX (UNSY ) UNIV SYDNEY.
XX
XX Abbott CA, Gorrell MD;
XX
XX WPI: 2002-454646/48.
XX N-PSDB: AAD38957.
XX
PT New dipeptidyl peptidase (DPP) peptidases, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
XX Claim 1; Page 70-74; 91pp; English.
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX polynucleotides encoding such proteins. The DPP peptidases are useful for
XX screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV (human immuno deficiency virus) infection. The present
XX sequence is an alternative version of murine DPP9 protein.
XX Note: This sequence is stated to be the same as that shown as
XX SEQ ID NO: 4 in the sequence listing of the specification. However these
XX sequences differ.
XX
SQ Sequence 869 AA;
Query Match 92.1%; Score 4279; DB 23; Length 869;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 792; Conservative 32; Mismatches 36; Indels 0; Gaps 0;
QY 4 TGTPTADRCDAATDDPAARFOYKHSMDGLRSIIHSGSRKYSGLIVNKAPHDFQVOKTD 63
DB 10 SGVSPVEQVAAQDMDDTAARFCVOYKHSMDGLRSIIHSGSRKSSGLIVSKAPHDFQVOKPD 69

```

```
QY 64 ESGPESHRLYYLGMPYCSRENSLLYSEIPKKVKREALLLSWKOMLDHFOATPHHGYSR 123
    |||
Db 70 ESGPESHRLYYLGMPYCSRENSLLYSEIPKKVKREALLLSWKOMLDHFOATPHHGYSR 129
QY 124 EEBLLRRKRRLGVGITSYDFHSGSLFLFOASNSLFHCRDGGKNGFMVSPMKPLEIKTQ 183
    |||
Db 130 EEBLLRRKRRLGVGITSYDFHSGSLFLFOASNSLFHCRDGGKNGFMVSPMKPLEIKTQ 189
QY 184 CSGPRMDPKICPADPAFFSFINSNDLWVANIENTGEERLFLFCHOGSLNVLDDPKSAGVAT 243
    |||
Db 190 CSGPRMDPKICPADPAFFSFINSNDLWVANIENTGEERLFLFCHOGSAGVLDNPKSAGVAT 249
QY 244 FVIOEEFDRFTGYWMCPTASWESSEGLKTLRLIYEEDSEVEVIHVPSPALBERKTDY 303
    |||
Db 250 FVIOEEFDRFTGYWMCPTASWESSEGLKTLRLIYEEDSEVEVIHVPSPALBERKTDY 309
QY 304 RYPTGSKNPKIALKLAELFOTDSOGKITVNOEKLVOPFSSLPFKVXIARAGTTRGKY 363
    |||
Db 310 RYPTGSKNPKIALKLAELFOTDSOGKITVNOEKLVOPFSSLPFKVXIARAGTTRGKY 369
QY 364 AMAAFLDROPQORQOLVLLPALFTIPAVESEAOQAARAARAVKNOPEVITYEETVNWIMV 423
    |||
Db 370 AMAAFLDROPQORQOLVLLPALFTIPAVESEAOQAARAARAVKNOPEVITYEETVNWIMV 429
QY 424 HDIFPPQSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSGQYDWSBPFGDEEFKCP1 483
    |||
Db 430 HDIFPPQSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSGQYDWSBPFGDEEFKCP1 489
QY 484 KEELALTSGEWEVYIARHGSKIWNNEETKLYVFOGKTPTLEHNLVYVSYEAGETVLT 543
    |||
Db 490 KEELALTSGEWEVYIARHGSKIWNNEETKLYVFOGKTPTLEHNLVYVSYEAGETVLT 549
QY 544 PGFSHSCSMSQNDPMFVSHYSVSTPCVHYKLSGPDDELHKOPRFMASMEASCP 603
    |||
Db 550 LGFSHSCSMSQSFDMFVSHYSVSTPCVHYKLSGPDDELHKOPRFMASMEASCP 609
QY 604 DYPPPELFHHTKSDVLYKMTKPRHALDQCKKHPTVLFYGGQVOLVNNSEFGIKYLR 663
    |||
Db 610 DYPPPELFHHTKSDVLYKMTKPRHALDQCKKHPTVLFYGGQVOLVNNSEFGIKYLR 669
QY 664 LNTLASIGYVAVVVDGSGSCORGREFGALKNMGQVEIEDQVGLTFVAKYGFIDLSR 723
    |||
Db 670 LNTLASIGYVAVVVDGSGSCORGREFGALKNMGQVEIEDQVGLTFVAKYGFIDLSR 729
QY 724 VAIHGMSYGFSLMGLIHKRQVFKVAIAGAPVTVMAYDTGTYERYMDYENNOGYEA 783
    |||
Db 730 VAIHGMSYGFSLMGLIHKRQVFKVAIAGAPVTVMAYDTGTYERYMDYENNOGYEA 789
QY 784 GSAVALHEKLPNEPNRLILHGFLEENVHFFHTNLFVSOLIRAGKPYOLOIYPERHSIR 843
    |||
Db 790 GSAVALHEKLPNEPNRLILHGFLEENVHFFHTNLFVSOLIRAGKPYOLOIYPERHSIR 849
QY 844 CPESGEHYEVTLLHFLQEYL 863
    |||
Db 850 CPESGEHYEVTLLHFLQEYL 869

RESULT 9
AAE23875
ID AAE23875 standard: Protein: 847 AA.
AC AAE23875;
XX
XX 23-SEP-2002 (first entry)
DE Murine dipeptidyl peptidase 9 (DPP9) protein.
XX
XX Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
KM human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
KM type II diabetes; antidiabetic; antiinflammatory; immunosuppressive;
XX antiviral; enzyme.
OS Mus sp.
```

```
XX XX
PN WO200234900-A1.
XX XX
XX 02-MAY-2002.
PD XX
PF 29-OCT-2001; 2001WO-A001388.
PR XX
PR 27-OCT-2000; 2000AU-0001078.
PA (UNSY ) UNIV SYDNEY.
PI Abbott CA, Gorrell MD;
XX
XX WPI: 2002-454646/48.
DR N-PSDB: AAD38311.
XX
PT New dipeptidyl peptidase (DPP) peptides, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
PS Claim 1; Fig 8; 91pp; English.
CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is an alternative version of murine DPP9 protein.
CC Note: This sequence is stated to be the same as that shown as
CC SEQ ID NO: 4 in the sequence listing of the specification. However these
CC sequences differ.
SQ Sequence 847 AA:
Query Match 88.9%; Score 4129; DB 23; Length 847;
Best Local Similarity 91.9%; Pred. No. 0;
Matches 764; Conservative 32; Mismatches 35; Indels 0; Gaps 0;
QY 4 TGTPTADRGDAATDDPAARFOVQKHSMDGLRSITIHGSRKYSGLIYKAPRDFQVQKTD 63
    |||
Db 10 SGVSPVQVAAGDMDDPAARFCVQKHSMDGLRSITIHGSRKYSGLIYKAPRDFQVQKTD 69
QY 64 ESGPESHRLYYLGMPYCSRENSLLYSEIPKKVKREALLLSWKOMLDHFOATPHHGYSR 123
    |||
Db 70 ESGPESHRLYYLGMPYCSRENSLLYSEIPKKVKREALLLSWKOMLDHFOATPHHGYSR 129
QY 124 EEBLLRRKRRLGVGITSYDFHSGSLFLFOASNSLFHCRDGGKNGFMVSPMKPLEIKTQ 183
    |||
Db 130 EEBLLRRKRRLGVGITSYDFHSGSLFLFOASNSLFHCRDGGKNGFMVSPMKPLEIKTQ 189
QY 184 CSGPRMDPKICPADPAFFSFINSNDLWVANIENTGEERLFLFCHOGSLNVLDDPKSAGVAT 243
    |||
Db 190 CSGPRMDPKICPADPAFFSFINSNDLWVANIENTGEERLFLFCHOGSAGVLDNPKSAGVAT 249
QY 244 FVIOEEFDRFTGYWMCPTASWESSEGLKTLRLIYEEDSEVEVIHVPSPALBERKTDY 303
    |||
Db 250 FVIOEEFDRFTGYWMCPTASWESSEGLKTLRLIYEEDSEVEVIHVPSPALBERKTDY 309
QY 304 RYPTGSKNPKIALKLAELFOTDSOGKITVNOEKLVOPFSSLPFKVXIARAGTTRGKY 363
    |||
Db 310 RYPTGSKNPKIALKLAELFOTDSOGKITVNOEKLVOPFSSLPFKVXIARAGTTRGKY 369
QY 364 AMAAFLDROPQORQOLVLLPALFTIPAVESEAOQAARAARAVKNOPEVITYEETVNWIMV 423
    |||
Db 370 AMAAFLDROPQORQOLVLLPALFTIPAVESEAOQAARAARAVKNOPEVITYEETVNWIMV 429
QY 424 HDIFPPQSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSGQYDWSBPFGDEEFKCP1 483
    |||
Db 430 HDIFPPQSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSGQYDWSBPFGDEEFKCP1 489
QY 484 KEELALTSGEWEVYIARHGSKIWNNEETKLYVFOGKTPTLEHNLVYVSYEAGETVLT 543
    |||
```

Db 490 KEEVALTSGEWEVLRSRHSKIMVNEQTKLYFGQTKDPTLEHHLTVVSEASGEIVRLTT 549
QY 544 PGFHSCHSCMSQNFDMFVSHYSSVSTPPCVNHYKLSGPRDDPLHKOPRWMASMEASCPP 603
Db 550 LGFHSCHSCMSQSFDMFVSHYSSVSTPPCVNHYKLSGPRDDPLHKOPRWMASMEASCPP 609
QY 604 DYVPEIFHFHTRSDVRLYGMITYKPHALOPGKKHPTVLFVYGGPOVOLVNNNSFKIKTLR 663
Db 610 DYVPEIFHFHTRADYOLKGMITYKPHALOPGKKHPTVLFVYGGPOVOLVNNNSFKIKTLR 669
QY 664 LNTLASLGYAVVVIDGRSCGRLFEFEGALKNOMQVEIEDQVEGLQFAEKGFIIDLSR 723
Db 670 LNTLASLGYAVVVIDGRSCGRLFEFEGALKNOMQVEIEDQVEGLQFAEKGFIIDLSR 729
QY 724 VAHGMSTGGFSLMGLLHKPOVFVALAGAPVYVMAVYDGYTRYVNDVPENNOHGIEA 783
Db 730 VAHGMSTGGFSLMGLLHKPOVFVALAGAPVYVMAVYDGYTRYVNDVPENNOHGIEA 789
QY 784 GSVALLHVEKLEPNERNLLILHGFLEENHFFHTNPLVSQILRACKPYOLOI 834
Db 790 GSVALLHVEKLEPNERNLLILHGFLEENHFFHTNPLVSQILRACKPYOLOI 840

RESULT 10
ABG61605
ID ABG61605 standard; Protein: 832 AA.

AC ABG61605:

DT 12-AUG-2002 (first entry)

DE Human DPRP-2 splice variant #4.

XX
KW Human: serine protease: dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.

XX Homo sapiens.
PN WO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR) FERRING BV.

PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI: 2002-444178/47.

DR N-PSDB: ABK83336.

XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain

PS Disclosure; Page 85-87; 113pp; English.

XX The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,

CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 832 AA:

Query Match 86.7%; Score 4026; DB 23; Length 832;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 757; Conservative 6; Mismatches 16; Indels 20; Gaps 3;

QY 1 MATGTPTPADGDAAATDPPARQOVOKHNSMDGLSIHSGRSKISGLIYNKAPHDFOVQ 60
Db 30 MATGTPTPADGDAAATDPPARQOVOKHNSMDGLSIHSGRSKISGLIYNKAPHDFOVQ 89
QY 61 KTDGSGPHSHRLYYLGMFYGSRNSLXSEIPKVKRKFAALLLSMKOMLDHFQATPHHGV 120
Db 90 KTDGSGPHSHRLYYLGMFYGSRNSLXSEIPKVKRKFAALLLSMKOMLDHFQATPHHGV 149
QY 121 YSREELLREKRLGVFGITSYDFHSESGLEFLQASNSLFHCRODGKNGFVSPKPLEI 180
Db 150 YSREELLREKRLGVFGITSYDFHSESGLEFLQASNSLFHCRODGKNGFVSPKPLEI 209
QY 181 KTQCGPRMDPKICPADAPFSTINSDLVWANTETGEBRRLTFCHQGLSNVLDPKSAG 240
Db 210 KTQCGPRMDPKICPADAPFSTINSDLVWANTETGEBRRLTFCHQGLSNVLDPKSAG 269
QY 241 VATFVIOEEDFRFTGYWMCPTASWEGSEGLTLRLIYEEVSESEVYHVSPPALEEKT 300
Db 270 VATFVIOEEDFRFTGYWMCPTASWEGSEGLTLRLIYEEVSESEVYHVSPPALEEKT 329
QY 301 DSYRPRGTGSKNPKIALKLAEFQYDTSQGIYSTQKELVOPFSSLFPRVEYIARAGWTRD 360
Db 330 DSYRPRGTGSKNPKIALKLAEFQYDTSQGIYSTQKELVOPFSSLFPRVEYIARAGWTRD 389
QY 361 GRYAMAMFLDRPQOMLOVLLPRAFIPTSTENEEDRLASARAVRPNVYVVEYVW 420
Db 390 GRYAMAMFLDRPQOMLOVLLPRAFIPTSTENEEDRLASARAVRPNVYVVEYVW 449
QY 421 INVHDIFFYFPOSSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSGQYDSEPFSPCEDEFK 480
Db 450 INVHDIFFYFPOSSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSGQYDSEPFSPCEDEFK 509
QY 481 CPIKEEIALTSGEWEVLARHSGKIWNNEETKLVYFGQTKDPLRLEHHLTVVSYEAGEIVR 540
Db 510 CPIKEEIALTSGEWEVLARHSGKIWNNEETKLVYFGQTKDPLRLEHHLTVVSYEAGEIVR 569
QY 541 LTTPEFHSCHSCMSQNFDMFVSHYSSVSTPPCVNHYKLSGPRDDPLHKOPRWMASMEAS 600
Db 570 LTTPEFHSCHSCMSQNFDMFVSHYSSVSTPPCVNHYKLSGPRDDPLHKOPRWMASMEAS 629
QY 601 CPPDVVPEIFHFHTRSDVRLYGMITYKPHALOPGKKHPTVLFVYGGPOVOLVNNNSFKIK 660
Db 630 CPPDVVPEIFHFHTRSDVRLYGMITYKPHALOPGKKHPTVLFVYGGPOVOLVNNNSFKIK 689
QY 661 YLRNLTASLGYAVVVIDGRSCGRLFEFEGALKNOMQVEIEDQVEGLQFAEKGFIID 720
Db 690 YLRNLTASLGYAVVVIDGRSCGRLFEFEGALKNOMQVEIEDQVEGLQFAEKGFIID 749
QY 721 LSRVAIHGMSTGGFSLMGLLHKPOVFVALAGAPVYVMAVYDGYTRYVNDVPENNOHG 780
Db 750 LSRVAIHGMSTGGFSLMGLLHKPOVFVALAGAPVYVMAVYDGYTRYVNDVPENNOHG 789
QY 781 YEAGSVALLHVEKLEPNERN 799
Db 790 RKRALFPHKLRPLRTPDSPR 808

RESULT 11
ABG61606
ID ABG61606 standard; Protein: 832 AA.

XX ABG61606;
 AC
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-2 splice variant #5.
 XX
 KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinnesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.
 XX
 OS Homo sapiens.
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Q1 S, Atkinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI; 2002-444178/47.
 DR N-PSDB: ABK83337.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 FT
 PS Disclosure; Page 88-90; 113pp; English.
 XX
 PS The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinasias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
 CC
 XX
 SQ Sequence 832 AA;
 Query Match 86.7%; Score 4026; DB 23; Length 832;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 757; Conservative 6; Mismatches 16; Indels 20; Gaps 3;
 QY 1 MATTGCTPAAADGDAADDDPAARQVQKHSMDGLRSITHSKRKISGLIVNKAAPDFQVQ 60
 Db 30 MATTGCTPAAADGDAADDDPAARQVQKHSMDGLRSITHSKRKISGLIVNKAAPDFQVQ 89
 QY 61 KTDSSGPHSHRLTYLGMVYSGRENSLSEIPKRYRKFAALLLSMKOMLDFHOATPHHGV 120
 Db 90 KTDSSGPHSHRLTYLGMVYSGRENSLSEIPKRYRKFAALLLSMKOMLDFHOATPHHGV 149
 QY 121 YSREELRLERKRLGVGITSYDFHSEGLFLFOASNSLFLFCRDGKNGFVNSPMKPLEI 180
 Db 150 YSREELRLERKRLGVGITSYDFHSEGLFLFOASNSLFLFCRDGKNGFVNSPMKPLEI 209
 QY 181 KTQCSGPMADPKICPADPAFFSFITNSDLVANIETGERRLTFCHGLSNVLDPPKSAG 240

Db 210 KTQCSGPMADPKICPADPAFFSFITNSDLVANIETGERRLTFCHGLSNVLDPPKSAG 269
 QY 241 VATFVIOEEDRFTGYMMCPASMESEGLKTLITVEEVDESVEYIHWSPALAEKRT 300
 Db 270 VATFVIOEEDRFTGYMMCPASMESEGLKTLITVEEVDESVEYIHWSPALAEKRT 329
 QY 301 DSYRYPRTSKNPKIALKAEFOYDSOGKIVSTOEKELVOPFSSLPKVEYIARAGWTRD 360
 Db 330 DSYRYPRTSKNPKIALKAEFOYDSOGKIVSTOEKELVOPFSSLPKVEYIARAGWTRD 389
 QY 361 GKIVAMAFLDPRQOMLOLVLLPPLFTPTBNEORLASARAVRNQPVYVEEVYVNW 420
 Db 390 GKIVAMAFLDPRQOMLOLVLLPPLFTPTBNEORLASARAVRNQPVYVEEVYVNW 449
 QY 421 INVHDFPPPOSSEGEDELCLFRANECKTGFCCHLYKTAVALKSGOWMSEPFSGEDEPK 480
 Db 450 INVHDFPPPOSSEGEDELCLFRANECKTGFCCHLYKTAVALKSGOWMSEPFSGEDEPK 509
 QY 481 CPIKEEIALTSGEWEVLARHSGKIMVNEETKLVYFOGKTDTPLSHLDYVSYEAGETIVR 540
 Db 510 CPIKEEIALTSGEWEVLARHSGKIMVNEETKLVYFOGKTDTPLSHLDYVSYEAGETIVR 569
 QY 541 LTPPGFSHSCMSQNFDMFVSHYSVSTPCVHYKLSGPDDELHKOPRFWASMEAS 600
 Db 570 LTPPGFSHSCMSQNFDMFVSHYSVSTPCVHYKLSGPDDELHKOPRFWASMEAS 629
 QY 601 CPDYPVPEPEIFHFHTRSDVRLYGMIVKPHALQPKKHPTVLFYVGGPOVOLVNSPFGIK 660
 Db 630 CPDYPVPEPEIFHFHTRSDVRLYGMIVKPHALQPKKHPTVLFYVGGPOVOLVNSPFGIK 689
 QY 661 YLRNLTLASLGAAVVVIDGSSCORGLFEGALKNOMQVEIEOVGALQFAEKYFTID 720
 Db 690 YLRNLTLASLGAAVVVIDGSSCORGLFEGALKNOMQVEIEOVGALQFAEKYFTID 749
 QY 721 LSRVAIHGMSYGGFLSLMGLIHKQVFKVATAGAPVYMAAYIDGYERYVADVENNQHG 780
 Db 750 LSRVAIHGMSYGGFLSLMGLIHKQVFKVATAGAPVYMAAYIDGYERYVADVENNQHG 789
 QY 781 YEAGSVALHVEKLPNEPMR 799
 Db 790 RKRALFPKRLPLPTDPSR 808
 RESULT 12
 ABG61609
 ID ABG61609 standard; Protein: 819 AA.
 XX
 AC ABG61609;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-2 splice variant #8.
 XX
 KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinnesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.
 XX
 OS Homo sapiens.
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.

Page 11

QY	601	CPEDYVPEPEIFHNHTKSDVLLGKMIKRNALDRGKKNPRVLYFYGGPRQVLYVNSPKGIK	660
Db	617	CPEDYVPEPEIFHNHTKSDVLLGKMIKRNALDRGKKNPRVLYFYGGPRQVLYVNSPKGIK	676
QY	661	YLRLNTLASLGYANVVIDGSGSCQGRIGREFGALKKNMGQVEIEDQVEGLQFAEKYGFID	720
Db	677	YLRLNTLASLGYAVVVIDGSGSCQGRIGREFGALKKNMGQVEIEDQVEGLQFAEKYGFID	736
QY	721	LSRVAIHGWSYGGFLSLMGLIHKPOYEKVAIACAPTYVMAYDTGYTERMDVPENNHG	780
Db	737	LSRVAIHGWSYGGFLSLMGLIHKPOYFK-----ADPLAYPR--LP-----G	776
QY	781	YEAGSVALLHVEKLPEEPNR	799
		: : :	
Db	777	RKRALFPNKLRLPTDPSR	795
RESULT 13			
ABG61610	ID	ABG61610 standard; Protein; 819 AA.	
XX			
AC		ABG61610;	
XX			
DT	12-AUG-2002	(first entry)	
XX			
DE		Human DPPP-2 splice variant #9.	
XX			
Human: serine protease; dipeptidyl peptidase IV-related protein; DPPP			

KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 XX
 OS Homo sapiens.
 XX
 FN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 ER 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junjen J;
 DR WPI; 2002-444178/47.
 XX
 DR N-PSDB; ABR83341.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 PT
 XX
 PS Disclosure; Page 101-103; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and

```

CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 819 AA;
Query Match 84.8%; Score 3938.5; DB 23; Length 819;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 744; Conservative 6; Mismatches 16; Indels 33; Gaps 4;

QY 1 MATGTPTADGDAATDDPARFOYOKHSMWGLRSITIHGSRKYSGLIVNKAPHPDFQV 60
   |||||||
Db 30 MATGTPTADRGDAADDDPARFOYOKHSMWGLRSITIHGSRKYSGLIVNKAPHPDFQV 89
   |||||||

QY 61 KTDESGPHSRLLYYLGMYPYSGRENSLLYSELPKVRKREALLLSWKQMLDFQATPHHG 120
   |||||||
Db 90 KTDESGPHSRLLYYLGMYPYSGRENSLLYSELPKVRKREALLLSWKQMLDFQATPHHG 149
   |||||||

QY 121 YSREELLREKRRLGVGIGITSYDFHSGGLFLFOASNSLFRGCGKNGFVSPMKPLEI 180
   |||||||
Db 150 YSREELLREKRRLGVGIGITSYDFHSGGLFLFOASNSLFRGCGKNGFVSPMKPLEI 209
   |||||||

QY 181 KTQCSGRMPDKICPADPAFFSFINNSDLWVANIEETGEERLTCFHQGLSNVLDPKSAG 240
   |||||||
Db 210 KTQCSGRMPDKICPADPAFFSFINNSDLWVANIEETGEERLTCFHQGLSNVLDPKSAG 269
   |||||||

QY 241 VAFVIOEBEDRFTGYWMCPTASWESSEGLKTLRLIYEEVDESEVEYIHPSPALBERKT 300
   |||||||
Db 270 VAFVIOEBEDRFTGYWMCPTASWESSEGLKTLRLIYEEVDESEVEYIHPSPALBERKT 329
   |||||||

QY 301 DSYVPTGSKNPKIAIKLAEFOFDSOGKIYVTOEKELVQFESSLFPKVEYIAAGMTRD 360
   |||||||
Db 330 DSYVPTGSKNPKIAIKLAEFOFDSOGKIYVTOEKELVQFESSLFPKVEYIAAGMTRD 389
   |||||||

QY 361 GKYAMAFLDPRQOQMLDVLPLPALFTPTENEQRIASARAVRNQPYVVEEVTNW 420
   |||||||
Db 390 GKYAMAFLDPRQOQMLDVLPLPALFTPTENEQRIASARAVRNQPYVVEEVTNW 449
   |||||||

QY 421 INVHDIYPPPOSGEDELCLFLRANECKTGFCILYKVTAVLKSGCYDMSSEPSGDEDFK 480
   |||||||
Db 450 INVHDIYPPPOSGEDELCLFLRANECKTGFCILYKVTAVLKSGCYDMSSEPSGDEDFK 509
   |||||||

QY 481 CPRIKEEALTSGEWEVLARHSGKITWNEERTKLYFOCTKDPLEHHLYVSYEAGETVR 540
   |||||||
Db 510 CPRIKEEALTSGEWEVLARHSGK-----GTKDPTLEHHLYVSYEAGETVR 556
   |||||||

QY 541 LTPPGFSHSCSMQNFDMFVSHYSVSTPCVHYKLSGPDDELHAKQRFEMASMEBAAS 600
   |||||||
Db 557 LTPPGFSHSCSMQNFDMFVSHYSVSTPCVHYKLSGPDDELHAKQRFEMASMEBAAS 616
   |||||||

QY 601 CPPDYVPELTFHFTSRDVRLLYGMITYRPHALQPKKHPVLVYVGGFQVOLVNNSEKGIK 660
   |||||||
Db 617 CPPDYVPELTFHFTSRDVRLLYGMITYRPHALQPKKHPVLVYVGGFQVOLVNNSEKGIK 676
   |||||||

QY 661 YLRANTLASIGYAVVVDGKSGCORGLRFGALKQNGOYEIEQVGLQFVARKYFID 720
   |||||||
Db 677 YLRANTLASIGYAVVVDGKSGCORGLRFGALKQNGOYEIEQVGLQFVARKYFID 736
   |||||||

QY 721 LSRVAIHGWSYGGFLSLMGLIHKPOVEKVAIAGAPVTWMAVDTGYTERVYDVEENNQG 780
   |||||||
Db 737 LSRVAIHGWSYGGFLSLMGLIHKPOVEK-----AQPLATYPR--LP-----G 776
   |||||||

QY 781 YEAGSVALHVEKLPENNR 799
   : : : : :
Db 777 RKRALFPFKLRLPTDPSR 795
   : : : : :

```

```

DE Human polypeptide SEQ ID NO 1869.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokine; thrombolytic; drug screening; arthritis; inflammation;
KM leukaemia.
XX
XX Homo sapiens.
XX
XX WO20015312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX
XX 25-APR-2000; 2000US-0552317.
XX
XX 09-JUL-2000; 2000US-0598042.
XX
XX 19-JUL-2000; 2000US-0620312.
XX
XX 03-AUG-2000; 2000US-0653450.
XX
XX 14-SEP-2000; 2000US-0662191.
XX
XX 19-OCT-2000; 2000US-0693036.
XX
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX
XX DR N-PDB; AAI57880.
XX
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries .
XX
XX PS Example 3; SEQ ID NO 1869; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA158642-AA162213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX SQ Sequence 737 AA;
Query Match 82.2%; Score 3817.5; DB 22; Length 737;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 721; Conservative 3; Mismatches 4; Indels 37; Gaps 4;

QY 108 MLDFQATPHHGYVSREELLREKRRLGVGIGITSYDFHSGGLFLFOASNSLFRGCGK 167
   |||||||
Db 1 MLDFQATPHHGYVSREELLREKRRLGVGIGITSYDFHSGGLFLFOASNSLFRGCGK 60
   |||||||

QY 168 NGFVWSPMKPLEIKTQCSGRMPDKICPADPAFFSFINNSDLWVANIEETGEERLTCFHQ 227
   |||||||
Db 61 NGFVWSPMKPLEIKTQCSGRMPDKICPADPAFFSFINNSDLWVANIEETGEERLTCFHQ 120
   |||||||

QY 228 GLSNVLDLDPKSAGYATFVIOEBEDRFTGYWMCPTASWESSEGLKTLRLIYEEVDESEVEY 287
   |||||||

```

Db	121	GLSNVLDDPKSAGATVIVIQEEDPRFNGYMMWCPTASMSSEBGLTKRLILYEEVDESEVEY	180
Qy	288	IHWDSPALEEKRTDSYRYPRTGSKNPKIALKLAEEQTDSQKIVSTQEKELVOPFSSLEP	347
Db	181	IHWSPSALEEKRTDSYRYPRTGSKNPKIALKLAEEQTDSQKIVSTQEKELVOPFSSLEP	240
Qy	348	KVEYIARAGWTRDCKYAMAMFLDRPQWLDLVLLPRALFIPSTENEGQRLASRAAPRVN	407
Db	241	KVEYIARAGWTRDCKYAMAMFLDRPQWLDLVLLPRALFIPSTENEGQRLASRAAPRVN	300
Qy	408	QPYVYVEEVTNWMJINVNHIFPYFQSGEDBELCLFRANECKTGCHLYKVTAYLKSOGYD	467
Db	301	QPYVYVEEVTNWMJINVNHIFPYFQSGEDBELCLFRANECKTGCHLYKVTAYLKSOGYD	360
Qy	468	WSEPFSPGEDFCFKPIKEEIALITSGEWEVYLARHGSKIWNVEETKLVYFGTKDTPLENNL	527
Db	361	WSEPFSPGEDFCFKPIKEEIALITSGEWEVYLARHGSKIWNVEETKLVYFGTKDTPLENNL	401
Qy	528	YVYSIEAAGELIVRLITPFGFSHSCMSQNFDMFVSHYSYSTPRCVNYUKISGPDDBPLK	587
Db	402	YVYSIEAAGELIVRLITPFGFSHSCMSQNFDMFVSHYSYSTPRCVNYUKISGPDDBPLK	461
Qy	588	QPRFASAMWEAASCPDPDVPEPIHFHTRSDPRLYGMITYKPRHALDPEKKHPTVLFVYGGP	647
Db	462	QPRFASAMWEAASCPDPDVPEPIHFHTRSDPRLYGMITYKPRHALDPEKKHPTVLFVYGGP	512
Qy	648	QVOLVNNSEFKIKYLRLLNTLASLYAAVVVIDGRSCSGRGIFEGALKNMQGOVEIDQVE	707
Db	513	QVOLVNNSEFKIKYLRLLNTLASLYAAVVVIDGRSCSGRGIFEGALKNMQGOVEIDQVE	572
Qy	708	GLQFYAEKYGFIIDSRAVHGMSTYGFSLDMLGRIHKPQVERVALIAGAPVTWMAVDTGYT	767
Db	573	GLQFYAEKYGFIIDSRAVHGMSTYGFSLDMLGRIHKPQVERVALIAGAPVTWMAVDTGYT	632
Qy	768	ERYMDVPRNNNGHYEAGSVALHVEKLENEPRRLILHGFIDENVHFNHTFVLSOLIRAG	827
Db	633	ERYMDVPRNNNGHYEAGSVALHVEKLENEPRRLILHGFIDENVHFNHTFVLSOLIRAG	692
Qy	828	KPYQL-----QIYPNEHRSIRCPESGEGHYEVTLLHFLOEYL	863
Db	693	KPYQLQVALPPVSPQIYPNEHRSIRCPESGEGHYEVTLLHFLOEYL	737
RESULT 15			
AAM40510			
ID	AAM40510	standard; Protein; 683 AA.	
XX	AAM40510;		
AC			
XX			
DT	22-OCT-2001	(first entry)	
XX			
DE		Human polypeptide SEQ ID NO 5441.	
XX			
KW		Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer	
KW		peripheral nervous system; neuropathy; central nervous system; CNS;	
KW		Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;	
KW		amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW		chemokine; thrombolytic; drug screening; arthritis; inflammation;	
KW		Leukemia.	
OS		Homo sapiens.	
XX			
PN	WO200153312-A1.		
PD			
XX	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000WO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-052317.		
PR	09-JUL-2000; 2000US-0586042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAH36642-AAH42213) with neurotrophic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, Leukaemias and C.N.S. disorders.

Note: The sequence data for this patent did not form part of the printed specification.

Matches 668; Conservative 3; Mismatches 3; Indels 37; Gaps 4;

QY	162	CSGGKNGFVSPMKPLKTKQCSPPHMDPKICADAPAFSEFINNSLWAAINETGERR	221
Db	1	CRDGKNGFVSPMKPLKTKQCSPPHMDPKICADAPAFSEFINNSLWAAINETGERR	60
QY	222	LTFCQGLSNVLDPKSAGAVTFYIQEEDFRFTGYWMCPTASWEGSGKLTRLIYEVD	281
Db	61	LTFCQGLSNVLDPKSAGAVTFYIQEEDFRFTGYWMCPTASWEGSGKLTRLIYEVD	120
QY	282	ESEVYIVPSPALAEKRTDSYRIRPRGSKNPKIALKLAERQDTDSOGKIYSTOEKELVOP	341
Db	121	ESEVYIVPSPALAEKRTDSYRIRPRGSKNPKIALKLAERQDTDSOGKIYSTOEKELVOP	180
QY	342	FSSLPKVEYLARAGWTKDGYAAMFLDRQOOWLQVLLRPALFTSTENEBORLASAR	401
Db	181	FSSLPKVEYLARAGWTKDGYAAMFLDRQOOWLQVLLRPALFTSTENEBORLASAR	240
QY	402	AVPRVVOYUYVEEYTNWIVNHDIYFPPOSEBEDELCLPLRANECKTGCHLYKTVAYL	461
Db	241	AVPRVVOYUYVEEYTNWIVNHDIYFPPOSEBEDELCLPLRANECKTGCHLYKTVAYL	300
QY	462	KSGQVDSSEPPSPGDEKCPKIEKIALTISEWEVYLARHSGKIWNBEETKLYVFOGKTOT	521
Db	301	KSGQVDSSEPPSPGDEKCPKIEKIALTISEWEVYLARHSGKIWNBEETKLYVFOGKTOT	361
QY	522	PLEHNLVUYVEAEIYRLTPTGFSHSCSASQNDMFVSHYSVSVPYCNHYKLSGPD	581
Db	362	PLEHNLVUYVEAEIYRLTPTGFSHSCSASQNDMFVSHYSVSVPYCNHYKLSGPD	421
QY	582	DDPLHKOPFVWASMEASCPDDVYVPEIFHFHTRSVDRLYGMITYKDHALQCPKKHPTVL	641
Db	421	DDPLHKOPFVWASMEASCPDDVYVPEIFHFHTRSVDRLYGMITYKDHALQCPKKHPTVL	481

```

Qy 642 FVYGGPOVQVNNNSFKIKYLRNLTLASLGAVVVIDGRGSCQGRGLRREGALKKNOMGOVE 701
    |||||||
Db 453 FVYGGPOVQVNNNSFKIKYLRNLTLASLGAVVVIDGRGSCQGRGLRREGALKKNOMGOVE 512
    |||||||
Qy 702 IEDOVEGLOFVAEKYGFIDLSRVAIHGWSYGGFSLMGLIHKPOVFKVAIAGAPVTWMA 761
    |||||||
Db 513 IEDOVEGLOFVAEKYGFIDLSRVAIHGWSYGGFSLMGLIHKPOVFKVAIAGAPVTWMA 572
    |||||||
Qy 762 YDTGYTERYMDVPENNNOHGYEAGSVLALHVEKLPNEPNRLLILHGFLEDNVHFFHTNFLVS 821
    |||||||
Db 573 YDTGYTERYMDVPENNNOHGYEAGSVLALHVEKLPNEPNRLLILHGFLEDNVHFFHTNFLVS 632
    |||||||
Qy 822 QLIRAGKPYQL-----QIYPNERHSIRCPESGEHYEVTLLHFLQEYL 863
    |||||||
Db 633 QLIRAGKPYQLQVALPVPSPQIYPNERHSIRCPESGEHYEVTLLHFLQEYL 683
    |||||||

```

Search completed: December 12, 2002, 12:04:24
 Job time : 37 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 08:31:42 : Search time 2383 Seconds
(without alignments)
17785.812 Million cell updates/sec

Title: US-09-976-674-4

Perfect score: 2617

Sequence: 1 caagctacatgacgccaccca.....tctgagcgccgcgagatccg 2617

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estbda:*
2: em_esthum:*
3: em_estlin:*
4: em_estlmv:*
5: em_estlmv:*
6: em_estlpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl1:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_estl3:*
13: gb_estl4:*
14: gb_estl5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mam:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	926	35.4	3143	11	AK016546 Mus muscu
2	897.2	34.3	1137	13	BM553230 AGENCOURT
3	890.8	34.0	1043	14	BQ051605 AGENCOURT
4	844.4	32.3	921	14	BQ877413 AGENCOURT
5	843.8	32.2	932	14	BQ684956 AGENCOURT
6	834	31.9	982	14	BQ678015 AGENCOURT

7	805.2	30.8	880	14	BQ897707	BQ897707 AGENCOURT
8	785.8	30.0	879	14 <td>BQ889360 <td>BQ889360 AGENCOURT</td> </td>	BQ889360 <td>BQ889360 AGENCOURT</td>	BQ889360 AGENCOURT
9	783.2	29.9	872	14 <td>BQ689588 <td>BQ689588 AGENCOURT</td> </td>	BQ689588 <td>BQ689588 AGENCOURT</td>	BQ689588 AGENCOURT
10	779.8	29.8	866	13 <td>BM461814 <td>BM461814 AGENCOURT</td> </td>	BM461814 <td>BM461814 AGENCOURT</td>	BM461814 AGENCOURT
11	770.6	29.4	910	14 <td>BQ949519 <td>BQ949519 AGENCOURT</td> </td>	BQ949519 <td>BQ949519 AGENCOURT</td>	BQ949519 AGENCOURT
12	767.8	29.3	902	14 <td>BQ642814 <td>BQ642814 AGENCOURT</td> </td>	BQ642814 <td>BQ642814 AGENCOURT</td>	BQ642814 AGENCOURT
13	767.4	29.3	1051	13 <td>BM915935 <td>BM915935 AGENCOURT</td> </td>	BM915935 <td>BM915935 AGENCOURT</td>	BM915935 AGENCOURT
14	757.8	29.0	787	13 <td>BI223668 <td>BI223668 AGENCOURT</td> </td>	BI223668 <td>BI223668 AGENCOURT</td>	BI223668 AGENCOURT
15	750	28.7	938	14 <td>BQ685659 <td>BQ685659 AGENCOURT</td> </td>	BQ685659 <td>BQ685659 AGENCOURT</td>	BQ685659 AGENCOURT
16	746.4	28.5	1030	13 <td>BM424091 <td>BM424091 AGENCOURT</td> </td>	BM424091 <td>BM424091 AGENCOURT</td>	BM424091 AGENCOURT
17	727.2	27.8	975	14 <td>BQ681307 <td>BQ681307 AGENCOURT</td> </td>	BQ681307 <td>BQ681307 AGENCOURT</td>	BQ681307 AGENCOURT
18	718.2	27.4	974	13 <td>BM423964 <td>BM423964 AGENCOURT</td> </td>	BM423964 <td>BM423964 AGENCOURT</td>	BM423964 AGENCOURT
19	696.4	26.6	747	13 <td>BI258117 <td>BI258117 AGENCOURT</td> </td>	BI258117 <td>BI258117 AGENCOURT</td>	BI258117 AGENCOURT
20	689.2	26.3	880	14 <td>BQ681942 <td>BQ681942 AGENCOURT</td> </td>	BQ681942 <td>BQ681942 AGENCOURT</td>	BQ681942 AGENCOURT
21	686	26.2	1223	14 <td>BM809492 <td>BM809492 AGENCOURT</td> </td>	BM809492 <td>BM809492 AGENCOURT</td>	BM809492 AGENCOURT
22	676	25.8	1195	13 <td>BM56123 <td>BM56123 AGENCOURT</td> </td>	BM56123 <td>BM56123 AGENCOURT</td>	BM56123 AGENCOURT
23	672.4	25.7	997	14 <td>BQ642151 <td>BQ642151 AGENCOURT</td> </td>	BQ642151 <td>BQ642151 AGENCOURT</td>	BQ642151 AGENCOURT
24	668	25.5	698	12 <td>BG108176 <td>BG108176 AGENCOURT</td> </td>	BG108176 <td>BG108176 AGENCOURT</td>	BG108176 AGENCOURT
25	661.8	25.3	822	12 <td>BG330712 <td>BG330712 AGENCOURT</td> </td>	BG330712 <td>BG330712 AGENCOURT</td>	BG330712 AGENCOURT
26	658.8	25.2	727	13 <td>BI855677 <td>BI855677 AGENCOURT</td> </td>	BI855677 <td>BI855677 AGENCOURT</td>	BI855677 AGENCOURT
27	656.2	25.1	828	12 <td>BG424090 <td>BG424090 AGENCOURT</td> </td>	BG424090 <td>BG424090 AGENCOURT</td>	BG424090 AGENCOURT
28	644.4	24.6	1032	12 <td>BG167854 <td>BG167854 AGENCOURT</td> </td>	BG167854 <td>BG167854 AGENCOURT</td>	BG167854 AGENCOURT
29	625.6	23.9	943	14 <td>BQ895618 <td>BQ895618 AGENCOURT</td> </td>	BQ895618 <td>BQ895618 AGENCOURT</td>	BQ895618 AGENCOURT
30	625.4	23.9	1080	12 <td>BF203783 <td>BF203783 AGENCOURT</td> </td>	BF203783 <td>BF203783 AGENCOURT</td>	BF203783 AGENCOURT
31	623.6	23.8	668	12 <td>BG290429 <td>BG290429 AGENCOURT</td> </td>	BG290429 <td>BG290429 AGENCOURT</td>	BG290429 AGENCOURT
32	622.4	23.8	1023	14 <td>BQ681703 <td>BQ681703 AGENCOURT</td> </td>	BQ681703 <td>BQ681703 AGENCOURT</td>	BQ681703 AGENCOURT
33	622	23.8	892	13 <td>BI160407 <td>BI160407 AGENCOURT</td> </td>	BI160407 <td>BI160407 AGENCOURT</td>	BI160407 AGENCOURT
34	619	23.7	821	13 <td>BI690774 <td>BI690774 AGENCOURT</td> </td>	BI690774 <td>BI690774 AGENCOURT</td>	BI690774 AGENCOURT
35	616.4	23.6	1088	13 <td>BM458946 <td>BM458946 AGENCOURT</td> </td>	BM458946 <td>BM458946 AGENCOURT</td>	BM458946 AGENCOURT
36	615.8	23.5	908	12 <td>BG767446 <td>BG767446 AGENCOURT</td> </td>	BG767446 <td>BG767446 AGENCOURT</td>	BG767446 AGENCOURT
37	611	23.3	931	12 <td>BG421210 <td>BG421210 AGENCOURT</td> </td>	BG421210 <td>BG421210 AGENCOURT</td>	BG421210 AGENCOURT
38	610.4	23.3	725	13 <td>BM018261 <td>BM018261 AGENCOURT</td> </td>	BM018261 <td>BM018261 AGENCOURT</td>	BM018261 AGENCOURT
39	597	22.8	653	13 <td>BI257144 <td>BI257144 AGENCOURT</td> </td>	BI257144 <td>BI257144 AGENCOURT</td>	BI257144 AGENCOURT
40	596.8	22.8	809	13 <td>BM045258 <td>BM045258 AGENCOURT</td> </td>	BM045258 <td>BM045258 AGENCOURT</td>	BM045258 AGENCOURT
41	595.6	22.8	744	10 <td>BE407965 <td>BE407965 AGENCOURT</td> </td>	BE407965 <td>BE407965 AGENCOURT</td>	BE407965 AGENCOURT
42	595.6	22.5	1000	13 <td>BM558851 <td>BM558851 AGENCOURT</td> </td>	BM558851 <td>BM558851 AGENCOURT</td>	BM558851 AGENCOURT
43	590	22.5	811	9 <td>AL529248 <td>AL529248 AGENCOURT</td> </td>	AL529248 <td>AL529248 AGENCOURT</td>	AL529248 AGENCOURT
44	588.4	22.5	611	10 <td>AM374284 <td>AM374284 AGENCOURT</td> </td>	AM374284 <td>AM374284 AGENCOURT</td>	AM374284 AGENCOURT
45	585.2	22.4	1077	13 <td>BM560972 <td>BM560972 AGENCOURT</td> </td>	BM560972 <td>BM560972 AGENCOURT</td>	BM560972 AGENCOURT

ALIGNMENTS

RESULT 1
AK016546
LOCUS 3143 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932434F09, homolog to DIPEPTIDYL PEPTIDASE 8, full insert sequence.
ACCESSION AK016546
VERSION AK016546.1 GI:12855334
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library
Clone:4932434F09.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL 99279253
MEDLINE 10349636
PUBMED 12
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, R., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20493374
PUBMED 11042159

Fri Dec 13 12:15:32 2002

3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carinci, P.,
Kono, H., Akiyama, Y., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishie, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujita, S., Inoue, K., Togawa, Y., Izawa, E., Matsubara, S., Kawai, J.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Mutsuura, S., Kawai, J.,
Okazaki, T., Muramatsu, M., Inoue, Y., Kiz, A. and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system—384-format
sequencing pipeline with 384 multipicillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE	4
AUTHORS	Kawai, J., Shingagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,

TITLE
JOURNAL

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

MEDLINE 21085660
 PUBMED 11217851
 REFERENCE 5 (bases 1 to 3143)

AUTHORS

Adachu, J., Alzawa, K., Akhilara, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hangaki, T., Hara, A., Hayatsu, N., Hill, D., Hitamoc, K., Hiraoka, T., Horii, F., Hume, D., Imolant, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, R., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shihm, L., Shbata, K., Shibata, Y., Shitagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejlina, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshinide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Chemical Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute: 1-1-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9232, Fax: 81-45-503-9216).

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGACAGCAGAGATCCAGACCTCTCTTTTITTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-

GAGAGGACATTTCTGGAGTAAATTAAATAAATCCCCCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was selected before cloning. Vector: a modified pluscript KS(+) after bulk excision from Lambda FLIC I. Cloning sites, 5' end: SalII; 3' end: BamBI. Host: DH10B.

FEATURES	SOURCE
Location/Qualifiers	
1. .3143	
/organism="Mus musculus"	
/strain="C57BL/6J"	
/db_xref="FANTOM_DB:4932434F09"	
/db_xref="MGI:MGI:1907554"	
/db_xref="taxon:10090"	
/clone="4932434F09"	
/sex="male"	
/tissue_type="testis"	
/clone_lib="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="adult"	
354. .3005	
CDS	

/note="data source:SPTR, source key:Q9HBM5, evidence:ISS homolog to DIPEPTIDYL PEPTIDASE 8

```

putative"
/codon_start=-1
/protein_id="Bab30295.1"
/db_xref="GI:12855335"
/db_xref="MISC:MGI:1921638"
/translation="MAAAMEEEDLGVIEPTACFEGNGSGSDPRKLEPYRYVRYMS
QKLKLLADLTRKYGCMYMAKAPRHPDPMFYKTRDPRPHSDRYTYLMSGENRENTLPFSE
IPTITNAAVALMLSMKPLDIDFOATIDYDLSSEELLRKRTIGVGLIAYDIAFHGS
GTLPLFAGSGITVHIDKGGPRGPTQDPRPLVLTSCPNIMDKRLCPADWDIAFHGS
NDWISLNLRYLERERRITTYVNEELANMEEDRSAGVATFVJQEGEDFRSGCMYKPMQER
TPGSGKILRLVEENDESEVEIIVHTSPMLTEEDRSFRYPKGTANRKYATFKMSEIV
VDAAAGIIVDIDKELVQPEELIEFGVEYVIRAAQMPFGKIAHMLIDRSPHQIOLV
SPFLPIVEDDADNRDLRLESDVSPPLIYETETDIWHIDIDHFVQTHQDELE
FPIASECKTEGRHLRYKITSLLKSKYKSSGGAPADSEFCPLKEETITSGMEWYG
RHSQNIWVEDARKLYVEPTGDKDPLEHLELVTVTGANVEVLRDITGYSNCSLHC
DFEITSYNNQKNPHCYSLYKSSLEDDEDPYKATKEPATLILDSAGPLPDYPTPEFSK
STTGPIYLLMYKRYHPOLOGKRYPLVFIYGGSGOOLNANREGVYVAFRPLNTLASIG
VYVYIYDNRSGCHRGKLFEGCAFYKMKQDIEIDDDVEGLQIYASQYDPLIDRVTIRHGS
YQGLYSMLALMQRSDLFRVATACAPVWLTFTGTERTWGAHPDQDNEQSTYLGSAW
OAGKFPSEPRKRLILHGFEDENHFAVHTSLTSLFVLVAKGRPYDLOIPRHSIRVPE
SGEYHEHLNLTYYGNSIRIALAKYI"
894 a      688 c      773 g      788 t

```

Query Match	Score	DB 11:	Length
35.4%	926	3143	

Best Local Similarity	61.08;	Pred. No. 7.3e-192;
Matches 1540; Conservative	0;	Mismatches 980; Indels 6; Gaps 2;

80 TTCCAGGTGCAAGACACTCGTGGGACGGGCTCCGGACATCATTCACGGGACCCGCAAG 139

Db 459 TTTTATGTCGACGGTATTCCTTGAGTCAGCTGAAAAAGCTGCTTGCTGATACACGAAAA 518

140 TACTTCGGGCTCATTGTTCACACAAGCGGCCCCACGACTTCCAGTTTGTGCAGAGACCGAT 199

Db 519 TACCACGGCTACATGATGGCTAAGCGGCCACATGACTTTATGTTTCTTAAGAGGACGGAT 578

200 GAGTCTGGGCCCCACTCCACCCGCTCTACTACCTGGGAATGCCATATGGCAGCCGAGAG 259

Db 579 CCACATAGGCTCAGACAGGGTCTATTACCTTGCCATGTCTGGTGAGACAAGAGAA 638

260 AACCTCCCTCCTCTACTCTGAGATTCCCAAGAAAGGTCCCGAAAGAGGCTCTGCCTGCCTCG 319

Db 639 AATACTATTTTATTCGAAATCCCTAAACCATCAACAGAGCAGCAGTCCTTATATGCTT 698

320 TCCGTGGAGCAGATGCTGGATCATTTCCAGGCCACGCCCCACCATGGGGTCTACTCTCGG 379

Db 699 TCCGGAAGCCCCCTTTGGATCTTTTCAGCGCAGACTAGACTATGGGATGTTCCGA 758

380 GAGGAGGAGCTGCTGAGGAGGCGGAAACGCCCTGGGGGCTTCCGCATCACCCTCCTACGAC 439

Db 759 GAGGAGAGCTACTCAGAGGAAGGCGCATTTGGAACCGTGGGATCCGACGTTATGAT 818

440 TTCACACAGCGAGAGTGGCTCTTTCCCTTTCCACAGGCCACACACAGCTCTTTCCACTGTCG 499

Db	819	TACCA	CCAGGAAGCGGACATTCCTGTTTCAAGCTGGGAGCGGCAATTACCATATCAAA	878
Qy	500	GACGCGCGCAAGAACGGCTTCATGATGTCGCCCTATGAAACCGCTTGAAATCAAGACCGAG	559	
Db	879	GATGAGAGACACATGATATTTACACAACAGCCTTTGGCGGCCCAATTTAGTGGAAACTRAGT	938	
Qy	560	TGCTCAGAGGCGCCGAGATGAGACCCCAAAATCTGCCCTGCCGACCTTGCCTCTCTC	619	
Db	939	TGTCCTCAATATAGGAATGGAATCCAAAATTTATGCCCTGTGATCCAGACTGGATAGCTTTC	998	
Qy	620	ATCAATTAACAGCGACCTGTGGGTGGCCAACTTCGAGACAGGAGAGACGGCGGCTGACC	679	
Db	999	ATTCAACGCAATGATATTTGGATATCAAACTCTGTACACGGGAAAGACGGAGATCCACA	1058	
Qy	680	TTTCGCCCAAGAGTTTATCATATGTCAGAGTAGACCCCAATCTGCGGGTGTGGCCAC	739	
Db	1059	TACGTACACAAATGAGCTAGCCAACTGGAAGAGATCCAGATCAGCTGGTGTGGCAC	1118	
Qy	740	TTTCGTATACAGAGAAAGATTGCAACCGCTTCACTGGGTACTGGTGTGCCACAGCCTCC	799	
Db	1119	TTTCTCTTCAAGAAAGAAATTTGACACATACTCTGGCTACTGGTGTGCCCAAGCGSAA	1178	
Qy	800	TGGGAAGCTTCAGAGGCGCTCAAGACGCTGCGAATCTGTATAGAGAAATGATGATGCC	859	
Db	1179	AG---AACTCCTGTGTGTGTATAAAATTTCTAGATTTCTCTATGAAGAAATGATGAACT	1235	
Qy	860	GAGGTGGAGGCAATCAACGTCCTCCCGGGCTAAGAAAGAAAGAACAGGACTCTGAT	919	
Db	1236	GAGGTGAGATTTATTCATGTTATGCTCCCACTGTTGGAAACAGGAGGCGAGATTCCTT	1295	
Qy	920	CGGTACCCCAAGAGACAGGACAGCAAGAAATCCCAAGATTGCTTGAACCTGGCTAGTCCAG	979	
Db	1296	CGTTATCCCAAAAGGACACAGCAAAACCAAGAGTCATTTACAATGTCCGAGATTGTT	1355	
Qy	980	ACTACACGCGAGGGCAAGATCGTCTGCAACCCAGAGAGAGAGAGCTGTGCACGCTTCAGC	1039	
Db	1356	GTTATCTCTCAGAGAGGATATAGATATGTCATAGATAAAGAACTGGTTCAACTTTTGAG	1415	
Qy	1040	TCGCTGTTCCCAAGGAGTGATCAATCGCAGAGGCCGGGTGACCCGGGATGGCAAAATAC	1099	
Db	1416	ATTTCTGTTTGAAGGAGATTGATATATTTGCCAGAGCCGGATGGAGCTCCAGAGGAAATAT	1475	
Qy	1100	GCTTGCGCATGTTCTCGACCGCGGCCCAAGCAGGTGGCTTCACCTGTCTCTCCGCCG	1159	
Db	1476	GCTGTGTCATCTACTAGTACAGACGCTGCCAGACTCAACCTGCATATGTTCTGTCTCCCT	1535	
Qy	1160	GCCCTGTTCAATCCGAGCACAGAGAAATGAGAGACAGCGGCTTAGCTTGCCAGAGCTGTC	1219	
Db	1536	GAGTATTAATCAATCCAGTAGAAGATGATGTCATGAGACAGACAGACTTATAGAGTCAGTT	1595	
Qy	1220	CCCAAGAAATGCAAGCCGCTATGTGTGTACGAGAGAGTACCACAGCTGTGATCAATGTT	1279	
Db	1356	CTGACTCTGTGACACCACGTGATCATCTATGAAGAAACAACAGCATGTGATTAATATTC	1655	
Qy	1280	CATGACATCTTCTATCCCTTCCCCCATATGAGAGGAGAGAGAGACTCTGCTTCTCCGC	1339	
Db	1656	CAGCATATTTTTCATGTTTTTCTCTCAACTCA--TGAAGATGAATTTGAGTTATTTTT	1712	
Qy	1340	GCCAAATATGCAAGACCGGCTTCTCCATTTGTACAAAGTCACGCGGTTTTAAATCC	1399	
Db	1713	GCTCTGAAATGCAAAACAGGTTTTCTCATCTGTATAAATACATTCATTTTAAAGAG	1772	
Qy	1400	CAGGGCTACGATTTGAGATGAGCCCTTCAGCCCGCGGGAAGATGAATTTAAGTCCCATTT	1459	
Db	1773	AGCAAAATTAACGGGTCACTGTGTGACATCACTTGCCCCCAAGGATTTTCAAGTGTCTATTC	1832	
Qy	1460	AAGGAAGAATGCTCTGACACAGCGGTGATGGAGAGTTTGGCCAGACAGGCTCCAG	1519	
Db	1833	AAAGAAAGAAATTAACAATTTACCACTGGTGAATGGGAAGTACTTGGCCGCAATGATCTAAT	1882	
Qy	1520	ATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGCGCACCAAGACAGCGCGCTG	1579	

D	b	1893	ATTCGGGTGATGAGACCGAAGAAAGCTGGTGTACTTTGAAGGACACCAAGAACTTCCTTTGG	1952
Q	y	1580	GAGCAACACCCCTCTACGTGTACAGCTATGAGGGCGGGGAGATGCTACGCTCACAG	1639
D	b	1953	GAGCATACACCTGTACGTGAGACAGTATAGCAAAACCTGGAGAAAGTGGTAGCGCTGACTAC	2012
Q	y	1640	CCCGGCTTCTCCATAGCTGCTCCATGAGCCAGAACCTTGACATGTTGTCAGCCACTAC	1699
D	b	2013	CGTGGCTACTACACTCTGCTGCTCCACAGCGGCGCATTTGACTCTCTTAATGAATGATAC	2072
Q	y	1700	AGCAGCGGAGACACGCCGCTCGTGTACAGTATACAACTGAGGGCCCCGACAGACAC	1759
D	b	2073	AGCAACCAAGAAATCCACACTGTGTGTCTCTTACAAACTGTCAAGTCTGTAGATATAC	2132
Q	y	1760	CCCCGACAGCAGCCGCCCTCTGTGGGTAGCATGATGAGGACAGCAGCTGCCCCCG	1819
D	b	2193	GACTACACCCCTCCAGAAATTTTGTCTTTGAAGATACATAGATTTACACTGTATGGA	2252
Q	y	1880	ATGATCTACAAGCCCCACGCCCTTGACAGCAGGAGAGACGCCACCGTCTCTTTGTA	1939
D	b	2253	ATGTTGTATAAAGCCTCATGTAGACTACAACCTGGAAAGAAATACCCACTGTGTATTTCATA	2312
Q	y	1940	TATGGAGGCCCCCAGGAGGACGTGGTGAATAACTCCTTCAAAAGGCAATCAAGTACTTGGG	1999
D	b	2313	TATGGTGTCCCCAGGAGGACGTGTGTAAACATCGGTTTAAAGGAGTCAAGTATTTCCG	2372
Q	y	2000	CTCAACACACTGGGCTCCCTCGGGCTACAGCCGCTGTTGTATTTAGAGGCGCGGCTCTGT	2059
D	b	2373	CTGAACACCCGTGGGCTCCCTGGGTATATGTGTGTGTGTATAGACACAGGSGATCCGT	2432
Q	y	2060	CACGAGAGGCTTCGGTTTCGAAGGGGCCCTGAAAAACCAATGGCCAGGTGAGATCGAG	2119
D	b	2433	CACCGAGACTTAATTTTGAAGCGGCTTTAAATATATAAATGGGTCAATATGAAATTCAT	2492
Q	y	2120	GACCAAGGAGGGGCTGCAAGTTCGTGTGGCCGAGAAATATAGGCTCATGTACCTGTAGCGGA	2179
D	b	2493	GATCAAGTGAAGAGACTCCAGTACCTAGCTACATGTAGACTCTCATTTGACTGTGATGGA	2552
Q	y	2180	GTTCGATCCATGGCTGGTCTACGGGGGCTTCTCTGCTCATGTGGGCTACTTCCACAG	2239
D	b	2553	GTGGGCAATCCACGGCTGGTCTATATGTGGGTACTCTCCCTGATGGCAATTAATGACAGG	2612
Q	y	2240	CCCCAGGTTCGAAGGTGGCCATTCGGGGTCCCGGCTACCGTCTGATGGGCTACAGAC	2299
D	b	2613	TCGGATATCTCTCCGGTGGCTATATGTGTGGGGCCCCAGTACACCTGTGATCTTCTATGAT	2672
Q	y	2300	ACAGGATACATGAGGCTCATATGAGACGTCCCGAGAACAAACACAGCAGGTATATAGGCG	2359
D	b	2673	ACAGGATACAGGGAGCCCTATATGGGTCACTCGACCAAAATGAACAGGGGCTACTACTTA	2732
Q	y	2360	GGTTCGCGGCTCGACAGTGGAGAAAGCTGCCAATAGACCCACCGCTTGCTTATCTC	2419
D	b	2793	GGATCTGTGGSCATGCAAGGAGGAAGATTCCTCCAGAAACCAACCGGTATCTCTCTTG	2792
Q	y	2420	CACGGCTCTCTGAGCAAAAGCTGTACATTTTTCACACACAACACTCTCTGTCCCACTG	2479
D	b	2793	CATGGTCTTGTGATGTAGAAATGTTCACCTTTGCACACACAGATATATGTGCTAGTTTATA	2852
Q	y	2480	ATCCGAGAGGGAACCTTACACAGCTTCAGATCTTACCACAGAGAGACACAGTATTTGCG	2539
D	b	2853	GTGAGGGCTGGAAAGCCATATGACTTATCAACATCTATCTCCAGAGAGGGACAGCATCGA	2912
Q	y	2540	TGCCCCAGGTGGGCGAGCACTATGAAGTACAGTCTGCTCACTTTCTATAGGAATAACTC	2599
D	b	2913	GTCTCTGAGTCTGGAGAACACTATGAACCTGACCTGCTCCACTTACCTTCAGAGAACCTT	2972
Q	y	2600	TGAGCG 2605	
D	b	2973	GGATCG 2978	

RESULT 2	BM553230	1137 bp	mRNA	linear	EST 20-FEB-2002
LOCUS	BM553230				
DEFINITION	AGNCOURT.6572641 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5467115				
ACCESSION	BM553230				
VERSION	BM553230.1	GI:18791777			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 1137)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbds@remail.nih.gov Tissue Procurement: DCTD/DTP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov plate: L1CM1968 row: j column: 12 High quality sequence stop: 672.				
FEATURES	location/Qualifiers				
source	1..1137				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone_image="5467115"				
	/clone_id="NIH_MGC_41"				
	/tissue_type="amelanotic melanoma, cell line"				
	/lab_host="DH10B (phage-resistant)"				
	/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCCAGAG(g). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."				
BASE COUNT	253 a 346 c 320 g 216 t				2 others
ORIGIN					
Query Match	34.3%; Score 897.2; DB 13; Length 1137;				
Best Local Similarity	98.5%; Pred. No. 9.4e-186;				
Matches 947;	Conservative 0; Mismatches 9; Indels 5; Gaps 4;				
QY	348	AGGCCAGCCCCCACCATGTTGGGCTCTACTCTGCGGAGGAGAGCTGCTGAGGGAGCGGAAC	407		
Db	1	AGGCCAGCCCCCACCATGTTGGGCTCTACTCTGCGGAGGAGAGCTGCTGAGGGAGCGGAAC	60		
QY	408	GCTTGGGGGCTCTGGGATCACTCCTCAAGCTTCCACAGGAGAGTGCGCTTCTCTCT	467		
Db	61	GCTTGGGGGCTCTGGGATCACTCCTCAAGCTTCCACAGGAGAGTGCGCTTCTCTCT	120		
QY	468	TTCAGGCCAGCAACAGCTCTTCCATCTGTCGCGAGCGGCGGAAGAAGCGCTTCATGTGT	527		
Db	121	TTCAGGCCAGCAACAGCTCTTCCATCTGTCGCGAGCGGCGGAAGAAGCGCTTCATGTGT	180		
QY	528	CCCCATGAACCGGTGGAAATCAAGACCCAGTGTCAAGGGCCCCGGATGGAAACCCCAAAA	587		
Db	181	CCCCATGAACCGGTGGAAATCAAGACCCAGTGTCTCAGGGCCCCGGATGGAAACCCCAAAA	240		
QY	588	TCTGCTGCTGCGGACCT	647		
Db	241	TCTGCTGCTGCGGACCT	300		
QY	648	ACATTCGAGACAGCGCGGAGGCGCGGCTGACCTTCTGCGCCACCAAGTTTATCAATGTCC	707		

Db	301	ACATGAGACAGAGGAGGCGGGTGACCTCTTGCCACCAGATTATCAATGTCC	360
QY	708	TGGATGACCCCAAGTCTGCGGGTGTGGCCACCTTGTATACAGAAAGTTCCAGCCGT	767
Db	361	TGGATGACCCCAAGTCTGCGGGTGTGGCCACCTTGTATACAGAAAGTTCCAGCCGT	420
QY	768	TCACGTGGGTACGAGTGAGTGCCTCACAGCCTCTTGGAAGTTCAAGAGGCCCTCAAGAGC	827
Db	421	TCACGTGGGTACGAGTGAGTGCCTCACAGCCTCTTGGAAGTTCAAGAGGCCCTCAAGAGC	480
QY	828	TGCGAATCCTGTATGAGGAAGTGCATGATGATCCGAGGTGAGATCATTCACGTCCCTCTC	887
Db	481	TGCGAATCCTGTATGAGGAAGTGCATGATGATCCGAGGTGAGATCATTCACGTCCCTCTC	540
QY	888	CTGGCCTGGAAGAAAGAAAGCAGCTGATATGCTTACCACCGAGCAGGACCAAGATC	947
Db	541	CTGGCCTGGAAGAAAGAAAGCAGCTGATATGCTTACCACCGAGCAGGACCAAGATC	600
QY	948	CCAAATTTGCTTGAATCTGAGTGCATGATCCAGACTACAGCAGCAGGACCAAGATCTCTCGA	1007
Db	601	CCAAATTTGCTTGAATCTGAGTGCATGATCCAGACTACAGCAGCAGGACCAAGATCTCTCGA	660
QY	1008	CCCAGAGAAAGAGAGCTGTGTGCAGCCCTTCAGCTCTGCTTCTCCCAAGTGGAGTACATCG	1067
Db	661	CCCAGAGAAAGAGAGCTGTGTGCAGCCCTTCAGCTCTGCTTCTCCCAAGTGGAGTACATCG	720
QY	1068	CCAGGCGCGGGGTGAGCCCGGAGTGGCAATACGCTGGGACATCTTCTCTGACCGGCCCC	1127
Db	721	CCAGGCGCGGGGTGAGCCCGGAGTGGCAATACGCTGGGACATCTTCTCTGACCGGCCCC	780
QY	1128	AGCACTGGCTCAGCTGCTCTCTTCCCCCGGCGCTTTCATCCCGACAGAGAATG	1187
Db	781	AGCACTGGCTCAGCTGCTCTCTTCCCCCGGCGCTTTCATCCCGACAGAGAATG	839
QY	1188	AGGACACAGCGGCTACCTCTGTCACAGAGCTGTCGCCAGAAATGTCCAGCCGATGTGGATG	1247
Db	840	AGGACACAGCGGCTACCTCTGTCACAGAGCTGTCGCCAGAAATGTCCAGCCGATGTGGATG	899
QY	1248	ACGA-GGAGGTCCACACAGCTCT-GGATCAATGTTCATG--ACATCTTTCATCCCTTCCC	1303
Db	900	ACGAGGAGAGGTCCACACAGCTGTGGATCAATGTTCATGTGACATCTTCTATCCCTTCCC	959
QY	1304	C 1304	
Db	960	C 960	
RESULT 3			
LOCUS	B0051605	1043 bp	mrna EST_29-MAR-2002
DEFINITION	AGENCOURT_6866328 NIH_MGC_106 Homo sapiens cdna clone IMAGE:5932532		
ACCESSION	B0051605		
KEYWORDS	B0051605.1 GI:19810945		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 1043)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cga@bbs.fda.gov		
	Tissue Procurement: Dr. Daniel McVicar, DBS/NCI		
	CDNA Library Preparation: Rubin Laboratory		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		

Db	1	GCCTTCTCTCCTTCATC-ATACAGCAGCCTGTGGTGGCCAAATCAGAGACAGCCAG	59
QY	665	GAGGCGGCGCTGACCTTCTGCACCAAGTTTATCCAAATGCTGGATGACCCCAAGCT	724
Db	60	GAGGCGGCGCTGACCTTCTGCACCAAGTTTATCCAAATGCTGGATGACCCCAAGCT	119
QY	725	GCGGGTGTGGCCACCTTGTGTATACAGGAAGATTTCAGCCCTTCACTGGGTACTGTGG	784
Db	120	GCGGGTGTGGCCACCTTGTGTATACAGGAAGATTTCAGCCCTTCACTGGGTACTGTGG	179
QY	785	TGCCCCACAGCCTCTGGGAGGTTCCAGAGGCCCTCAGACGCTCCGAATCTGTATAG	844
Db	180	TGCCCCACAGCCTCTGGGAGGTTCCAGAGGCCCTCAGACGCTCCGAATCTGTATAG	239
QY	845	GAATTCATGATGCTCGAGAGTGGAGTCAATTCACGTCCTCTCCCTGAGGAAGAAG	904
Db	240	GAATTCATGATGCTCGAGAGTGGAGTCAATTCACGTCCTCTCCCTGAGGAAGAAG	299
QY	905	AAGACGACTGTATGCTGATCCAGAGCAGAGCAAGAAATCCCAAGATTGCTTGA	964
Db	300	AAGACGACTGTATGCTGATCCAGAGCAGAGCAAGAAATCCCAAGATTGCTTGA	359
QY	965	CTGGCTGATTCAGACTGACAGCAGAGGCAAGATCTCTGACCCAGAGAGAGCTG	1024
Db	360	CTGGCTGATTCAGACTGACAGCAGAGGCAAGATCTCTGACCCAGAGAGAGCTG	419
QY	1025	GTGAGCCCTTCACTGCTGTCCCGAAGGTGAGTACATCGCCAGAGGCGGGGTGAC	1084
Db	420	GTGAGCCCTTCACTGCTGTCCCGAAGGTGAGTACATCGCCAGAGGCGGGGTGAC	479
QY	1085	CGGATGGCAAAATACGCTGGGCAATGTTCTGAGCCGCGCCACAGCAGTGGCTCAGCTC	1144
Db	480	CGGATGGCAAAATACGCTGGGCAATGTTCTGAGCCGCGCCACAGCAGTGGCTCAGCTC	539
QY	1145	GTCCTCCCTCCCGGCGGCTTTCATCCCGACAGAGAAATGAGAGACAGCGCTACCC	1204
Db	540	GTCCTCCCTCCCGGCGGCTTTCATCCCGACAGAGAAATGAGAGACAGCGCTACCC	599
QY	1205	TCTCCAGAGCTGTCCCGAGAAATGTCAGCCGATGTGTGTAGAGAGAGTCAACAC	1264
Db	600	TCTCCAGAGCTGTCCCGAGAAATGTCAGCCGATGTGTGTAGAGAGAGTCAACAC	659
QY	1265	GTCGTGATCAATGTTTATGATCATCTTCTATCCCTTCCCAATCAGAGAGAGAGAG	1324
Db	660	GTCGTGATCAATGTTTATGATCATCTTCTATCCCTTCCCAATCAGAGAGAGAGAG	719
QY	1325	CTCGCTTCTCCCGGCAATGAATGCAAGCCGCTCTGCCATTTGTAAAGTCAACC	1384
Db	720	CTCGCTTCTCCCGGCAATGAATGCAAGCCGCTCTGCCATTTGTAAAGTCAACC	779
QY	1385	GCCGTTTTAAATCCCAAGGCTAGATTTGAGTGAAGCCCTTCAAGCCGCGGGAAGATGA	1443
Db	780	GCCGTTTTAAATCCCAAGGCTAGATTTGAGTGAAGCCCTTCAAGCCGCGGGAAGATGA	839
QY	1444	ATTTAAGTGGCCCATTAAGAAGATTGCTCTGACCAAGCGGTGA	1489
Db	840	ATTTAAGTGGCCCATTAAGAAGATTGCTCTGACCAAGCGGTGA	885
RESULT 5			
LOCUS	B0684956	932 bp	mRNA linear EST 15-JUL-2002
DEFINITION	AGENCOURT_8343602 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6250407		
ACCESSION	B0684956		
VERSION	B0684956.1	GI:21810272	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 932)		
AUTHORS	NIH-MGC	http://mgc.ncl.nih.gov/.	

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)									
JOURNAL	Unpublished (1999)									
COMMENT	Contact: Robert Strausberg, Ph.D. Email: gcrabs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution Information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.nih.gov Plate: LNCM293 High quality sequence stop: 710.									
FEATURES	Location/Qualifiers									
SOURCE	1..932 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6250407" /clone_1db="NIH_MGC_110" /tissue_type="ductal carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: pancreas; Vector: pOT87; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(5). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."									
BASE COUNT	209	a	265	c	269	g	188	t	1	others
ORIGIN										
Query Match	32.2%; Score 843.8; DB 14; Length 932;									
Best Local Similarity	96.9%; Pred. No. 4.2e-174;									
Matches 903; Conservative	0;	Mismatches	23;	Indels	6;	Gaps	4;			
QY	605	GCCTTCTTCTCCTTCATCATATACAGAGCCTGTGGGTGGCCAAATCGAGACAGCGAG	664							
Db	1	GCCTTCTTCTCCTTCATCATATACAGAGCCTGTGGGTGGCCAAATCGAGACAGCGAG	60							
QY	665	GAGGCGGCGCTGACCTTCTGCCACCAAGTTTATCCAAATGCTCGATGATGATGATGATG	724							
Db	61	GAGGCGGCGCTGACCTTCTGCCACCAAGTTTATCCAAATGCTCGATGATGATGATGATG	120							
QY	725	GCGGGTGTGGCCACCTTGTGTATGATGATGATGATGATGATGATGATGATGATGATG	784							
Db	121	GCGGGTGTGGCCACCTTGTGTATGATGATGATGATGATGATGATGATGATGATGATG	180							
QY	785	TGCCCAAGGCTCTGTGGAAAGTTTCAGAGGCGCTCAAGAGCTGCAATCTGTATGAG	844							
Db	181	TGCCCAAGGCTCTGTGGAAAGTTTCAGAGGCGCTCAAGAGCTGCAATCTGTATGAG	240							
QY	845	GAATTCATGATGCTCGAGAGTGGAGTCAATTCAGCTCCCTCTCTGGCTGAGGAAGAAG	904							
Db	241	GAATTCATGATGCTCGAGAGTGGAGTCAATTCAGCTCCCTCTCTGGCTGAGGAAGAAG	300							
QY	905	AAGACGACTGTATGCTGATCCAGAGCAGAGCAAGAAATCCCAAGATTGCTTGA	964							
Db	301	AAGACGACTGTATGCTGATCCAGAGCAGAGCAAGAAATCCCAAGATTGCTTGA	360							
QY	965	CTGGCTGATTCAGACTGACAGCCAGGCAAGATGCTTCTGACCCAGAGAGAGAGCTG	1024							
Db	361	CTGGCTGATTCAGACTGACAGCCAGGCAAGATGCTTCTGACCCAGAGAGAGAGCTG	420							
QY	1025	GTGAGCCCTTCACTGCTGTCCCGAAGGTGAGTACATCGCCAGAGGCGGGGTGAC	1084							
Db	421	GTGAGCCCTTCACTGCTGTCCCGAAGGTGAGTACATCGCCAGAGGCGGGGTGAC	480							
QY	1085	CGGATGGCAAAATACGCTGGGCAATGTTCTTGAACCGGCGCCAGAGTGGCTCAGCTC	1144							
Db	481	CGGATGGCAAAATACGCTGGGCAATGTTCTTGAACCGGCGCCAGAGTGGCTCAGCTC	540							
QY	1145	GTCTCTCTCCCGGCGGCTGTTCATCCCGAGCAGAGAAATGAGAGAGCGGCTAGCC	1204							

D6	541	GTCCTCCTCCCCCGGGCCCTTTCATGCCAGACACAAGAATTGAGACACC GGCTTAGCC	600
QY	1205	TCTGCACAAAGCTGTGCCAGGAATGTCCAGCCGATGTGTTGATGACGAGAGGTCAAC	1264
Locus	601	TCTGCACAAAGCTGTGCCAGGAATGTCCAGCCGATGTGTTGATGACGAGAGGTCAAC	660
OY	1265	GTCGTGATCAATGTTCATGACATCTTCTCATCCCTTCCCACATCAG-AGSAGAGACGA	1323
D6	661	GTCGTGATCAATGTTCATGACATCTTCTCATCCCTTCCCACATCAGAGGAGAGACGA	720
OY	1324	GCTCTGCTTCTCCGGCCCATGAATGCAACCGGCTTCTGCATTTGTCAAAGTCAC	1383
D6	721	GCTCTGCTTCTCCGGCCCATGAATGCAACCGGCTTCTGCATTTGTCAAAGTCAC	780
QY	1384	CGCCGTTTTAAATCCCAAGGGCTACGATTTGAGTGAGCCCTTCAGCCCGGGGAAATGA	1443
D6	781	CGCCGTTTT-AAATCCCAAGGGCTACGATTTGAGTGAGTA-CCTTCCGGCCCGGGGAAATGA	838
OY	1444	ATTAAAGTCCCATTTAAGAAAGATTTGCTCTGACCAAGCGGTGAA--TGGAGGTTTT	1500
D6	839	ATTAAAGTCCCATTTAAGAAAGATTTGCTCTGACCAAGCGGGAATAATGGGGAAGTTTT	898
OY	1501	GGCGAGCACAGGCTTCCAAGATCTGGGTCAATG	1532
D6	899	GGCGAGCACAGGCTTCCAATAATTGGGTCAATG	930
RESULT 6	B0678015	982 bp mRNA linear EST 15-JUL-2002	
LOCUS	B0678015		
DEFINITION	AGENCOUPT_8034317 NIH_MGC_112 Homo sapiens cdna clone IMAGE:6213986		
ACCESSION	B0678015		
VERSION	B0678015.1 GI:21790694		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: DCTD/DTP CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://Image.lnl.nih.gov Plate: LCM2380 row: J column: O3 High quality sequence stop: 683. Location/Qualifiers 1..982 /organism="Homo sapiens" /db_xref="taxon:9606" /cclone_image="IMAGE:6213986" /clone_id="NIH_MGC_112" /tissue_type="melanotic melanoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGGCGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."		
FEATURES	source		
BASE COUNT	202 a 310 c 290 g 179 t 1 others		
ORIGIN			

Query Match	31.98:	Score 834:	DB 14:	Length 982:
Best Local Similarity	96.78:	Pred. No. 6e-172:		
Matches 893:	Conservative 0:	Mismatches 26:	Indels 4:	Gaps 4:
Oy	1473	CTCTGACCAAGGCGTAAATGGGAGGTTTGGCGAGGACAGCGCTCCAAAGATCTGGGTCATATG	1532	
Db	1	CTCTGACCAAGCGGCGTAAATGGGAGGTTTGGCGAGGACAGCGCTCCAAAGATCTGGGTCATATG	60	
Oy	1533	AGGAGACCAAGCTGGTGTACTTCCAGGGCACCAAGACACGCCGCTGGAGCACCACTCT	1592	
Db	61	AGGAGACCAAGCTGGTGTACTTCCAGGGCACCAAGAGACACGCCGCTGGAGCACCACTCT	120	
Oy	1593	ACGTGTACGCTATGAGCGCGCGCGCGAGATCTACGCTTACCAACGCGCGCTTCTCC	1652	
Db	121	ACGTGTACGCTATGAGCGCGCGCGCGAGATCTACGCTTACCAACGCGCGCTTCTCC	180	
Oy	1653	ATAGTGTCTCATGAGCCAGAACTTCGACATGTTCTGTAGGCACATCAGCAGCGGACCA	1712	
Db	181	ATAGTGTCTCATGAGCCAGAACTTCGACATGTTCTGTAGGCACATCAGCAGCGGACCA	240	
Oy	1713	CGCGCGCTTGGGTGCAGCTTACAAAGCTGAGGGGCGCCGAGAGACGCCCTGCACAAGC	1772	
Db	241	CGCGCGCTTGGGTGCAGCTTACAAAGCTGAGGGGCGCCGAGAGACGCCCTGCACAAGC	300	
Oy	1773	AGCCCCGCTTCTGGGCTACGATGAGAGGACAGCAGCTGCCCGCCCGATTAATGTCTC	1832	
Db	301	AGCCCCGCTTCTGGGCTACGATGAGAGGACAGCAGCTGCCCGCCCGATTAATGTCTC	360	
Oy	1833	CAGAGATCTTCATTTCCACACAGCGCTGAGATGTGCGGCTCAGCGGATGATCTCAAGC	1892	
Db	361	CAGAGATCTTCATTTCCACACAGCGCTGAGATGTGCGGCTCAGCGGATGATCTCAAGC	420	
Oy	1893	CCACAGCCTTGACAGCCAGGAGAAAGAACCCACCCTCTCTTTGTAATGAGAGGCCCC	1952	
Db	421	CCACAGCCTTGAGCCAGGAGAAAGAACCCACCCTCTCTTTGTAATGAGAGGCCCC	480	
Oy	1953	AGGTCACTGCTGTAATTAATCTCTTCAAAAGCATTAATGCTGGCGCTCAACACACTGG	2012	
Db	481	AGGTCACTGCTGTAATTAATCTCTTCAAAAGCATTAATGCTGGCGCTCAACACACTGG	540	
Oy	2013	CCTCCCTGGGCTACGCGCGTGGTTGTGATTGAAGCGAGGGGCTCTGTACGAGAGGCTTC	2072	
Db	541	CCTCCCTGGGCTACGCGCGTGGTTGTGATTGAAGCGAGGGGCTCTGTACGAGAGGCTTC	600	
Oy	2073	GGTTGAAAGGGGCTGTAAAAACAATGGGCCAGGTGAGAGATCGAGGACAGGTGAGG	2132	
Db	601	GGTTGAAAGGGGCTGTAAAAACAATGGGGCAAGTGAAGTCAAGGACAGGTGAGG	660	
Oy	2133	GGCTCACTTCTGTGGCCGAGAAAGTATGGCTTCATGACTGAGCCGAGTGGCATCTCAT	2192	
Db	661	GGCTCACTTCTGTGGCCGAGAAAGTATGGCTTCATGACTGAGCCGAGTGGCATCTCAT	720	
Oy	2193	GCTGTCTTACAGGGGGCTCTCTGTGCTCATGGGCTAATCCACAAAGCCCGAGTGTCA	2252	
Db	721	GCTGTCTTACAGGGGGCTCTCTGTGCTCATGGGCTAATCC-CAAGCCCAAGNTGTCA	779	
Oy	2253	AGGTGCGCATGCGGGGTG-CCCCGTACACGCTGTGGATGGCTTACGACACAGGGTACT	2311	
Db	780	AGGTGCGCATGCGGGGTGCCCCCGGTACACGCTGTGGATGGCTTACGACACAGGGTACT	839	
Oy	2312	GAGC-GCTAACATGAGCGTCCCTGAGAACAAACAGCAGCGCTATGAGGGGGG-TTCGAGG	2369	
Db	840	GAGCGGCTACATGAGACGCTCTGAGAACAAACAGCAGCGCTTGAAGAGGGGGTTTCCGGGG	899	
Oy	2370	CCCTGCACGTGAGAGAGCTGCC 2392		
Db	900	CCCTGCACGTGAGAGAGAGCTGCC 922		


```

DEFINITION  AGENCOURT_8061873 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208407
5', mRNA sequence.
ACCESSION   B0897707
VERSION     B0897707.1 GI:22289721
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 880)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LNCM2366 row: a column: 16
            High quality sequence stop: 704.
FEATURES
    source
        1..880
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:6208407"
        /clone_1ib="NIH_MGC_110"
        /tissue_type="ductal carcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: pancreas; Vector: pOT7; Site_1: XhoI;
        Site_2: EcoRI; cDNA made by oligo-dT priming.
        Directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCGAGG(G). Library constructed by
        Ling Hong in the laboratory of Gerald M. Rubin (University
        of California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies).
        Note: this is a NIH_MGC library."
BASE COUNT  194 a 260 c 249 g 176 t 1 others
ORIGIN
Query Match 30.8%; Score 805.2; DB 14; Length 880;
Best Local Similarity 97.9%; Pred. No. 1.2e-165;
Matches 848; Conservative 0; Mismatches 13; Indels 5; Gaps 3;
QY 605 GCCTTCTTCCTTCATCATACAGAGGAGCTGTGGTGGCCAAATCGAGACAGGGGAG 664
    |||||||
Db 1 GCCTTCTTCCTTCATCATACAGAGGAGCTGTGGTGGCCAAATCGAGACAGGGGAG 60
QY 665 GACGCGGGGCTGACCTCTGCGACCAAGTTTATCCATGCTCGATGACCCCAAGTCT 724
    |||||||
Db 61 GACGCGGGGCTGACCTCTGCGACCAAGTTTATCCATGCTCGATGACCCCAAGTCT 120
QY 725 GCGGGTGTGGCCACCTTCGTCATACAGAAAGTTTCAGACCGCTTCACTGGGTACTGTGG 784
    |||||||
Db 121 GCGGGTGTGGCCACCTTCGTCATACAGAAAGTTTCAGACCGCTTCACTGGGTACTGTGG 180
QY 785 TGGCCCAACACCTTCCTGGGAAGTTTCAGAGGGCTCAAGACGCTGGCAATCTCTATGAG 844
    |||||||
Db 181 TGGCCCAACACCTTCCTGGGAAGTTTCAGAGGGCTCAAGACGCTGGCAATCTCTATGAG 240
QY 845 GAAGTCATAGTCCGAGGGAGGATTCACGCTCCCTCTCTGGGTAGAGAAAGG 904
    |||||||
Db 241 GAAGTCATAGTCCGAGGGAGGATTCACGCTCCCTCTCTGGGTAGAGAAAGG 300
QY 905 AAGACGAGCTCGATCGGTACCCAGAGACAGCAAGAAATCCCAAGATTGGCTTGAAG 964
    |||||||
Db 301 AAGACGAGCTCGATCGGTACCCAGAGACAGCAAGAAATCCCAAGATTGGCTTGAAG 360
QY 965 CTGGCTGAGTTCCAGACTGACGACGAGGCAAGATCGTCTCGACCCAGAGAAAGAGCTG 1024
    |||||||

```

```

Db 361 CTGGCTGAGTTCCAGACTGACAGCCAGGCGCAAGATCGTCTCGACCCAGAGAAAGAGCTG 420
QY 1025 GTGCAGCCCTTCACCTGCTGCTTCCCGAGAGTGAATACATCGCCAGAGCGCGGTGAGCC 1084
    |||||||
Db 421 GTGCAGCCCTTCACCTGCTGCTTCCCGAGAGTGAATACATCGCCAGAGCGCGGTGAGCC 480
QY 1085 CGGATGGAATATACGCTGTGGCCATGTCTCTGAGCCGCGCCAGAGTGGCTTCACCTC 1144
    |||||||
Db 481 CGGATGGAATATACGCTGTGGCCATGTCTCTGAGCCGCGCCAGAGTGGCTTCACCTC 540
QY 1145 GTCTCTCTCCCGCGCGCGCTTTCATCCCGAGACACAGAAATGAGAGCAGCGGCTGACC 1204
    |||||||
Db 541 GTCTCTCTCCCGCGCGCGCTTTCATCCCGAGACACAGAAATGAGAGCAGCGGCTGACC 600
QY 1205 TCTGCCAGACTGTCCCGAGAGATGTCCAGCCGATGTGTGATGAGAGAGTACCAAC 1264
    |||||||
Db 601 TCTGCCAGACTGTCCCGAGAGATGTCCAGCCGATGTGTGATGAGAGAGTACCAAC 660
QY 1265 GTCTGATCAATGTTCATGACATCTTATCCCTTCCCGCAATCAGAGAGAGAGAGAG 1324
    |||||||
Db 661 GTCTGATCAATGTTCATGACATCTTATCCCTTCCCGCAATCAGAGAGAGAGAGAGAG 720
QY 1325 CTCTGCTTCTCCCGCGCGCGCATGATGACAGACCGGCTTTCGCCATTTGTACAAAGTACC 1384
    |||||||
Db 721 CTCTGCTTCTCCCGCGCGCGCATGATGACAGACCGGCTTTCGCCATTTGTACAAAGTACC 780
QY 1385 G-CGGTTTAAATCCAGAGG--CTACGATTGAGTAGAGCC--TTACAGCCCGGAGAG 1439
    |||||||
Db 781 GCCCGTTTAAATCCAGAGGCGCTACGATTTGAGTAGAGCCCTTACAGCCCGGAGAG 840
QY 1440 ATGAATTTAAGTGGCCCATTAAGGAA 1465
    |||||||
Db 841 ATGAATTTAAGTGGCCCATTAAGGAA 866
RESULT 8
B0899360 879 bp mRNA linear EST 16-AUG-2002
LOCUS B0899360
DEFINITION AGENCOURT_8061873 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208551
5', mRNA sequence.
ACCESSION B0899360
VERSION B0899360.1 GI:22281374
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LNCM2366 row: g column: 16
            High quality sequence stop: 709.
FEATURES
    source
        1..879
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:6208551"
        /clone_1ib="NIH_MGC_110"
        /tissue_type="ductal carcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: pancreas; Vector: pOT7; Site_1: XhoI;
        Site_2: EcoRI; cDNA made by oligo-dT priming.
        Directionally cloned into EcoRI/XhoI sites using the

```


following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 194 a 256 c 252 g 177 t
ORIGIN

Query Match 30.0%; Score 785.8; DB 14; Length 879;

Best Local Similarity 99.6%; Pred. No. 2e-161;

Matches 798; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 605 GCCTCTCTCTCTCATATTAACAGCAGCTGTGTGGCCCAACATCGAGACAGCGGAG 664
|||||
Db 1 GCCTCTCTCTCTCATATTAACAGCAGCTGTGTGGCCCAACATCGAGACAGCGGAG 60
OY 665 GAGCGGGGCTGACCTTCTGCGACCAAGTTTATTCATATGCTTGATGACCCCAAGTCT 724
|||||
Db 61 GAGCGGGGCTGACCTTCTGCGACCAAGTTTATTCATATGCTTGATGACCCCAAGTCT 120
OY 725 GCGGGTGTGGCCACTTGTCTATACAGAAAGTTCGACCGCTTACTGGGTACTGTGG 784
|||||
Db 121 GCGGGTGTGGCCACTTGTCTATACAGAAAGTTCGACCGCTTACTGGGTACTGTGG 180
OY 785 TGCCCCACAGCCTCTCTGGGAAGTTCAGAGGGCTCAAGACGCTCGAATCCTGTATGAG 844
|||||
Db 181 TGCCCCACAGCCTCTCTGGGAAGTTCAGAGGGCTCAAGACGCTCGAATCCTGTATGAG 240
OY 845 GAAGTCGATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAGAACAAAGG 904
|||||
Db 241 GAAGTCGATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAGAACAAAGG 300
OY 905 AAGAGGAGCTGTATCGGTATACCCCAAGGACAGGACAGCAAGATCCCAAGTTCCCTTGA 964
|||||
Db 301 AAGAGGAGCTGTATCGGTATACCCCAAGGACAGGACAGCAAGATCCCAAGTTCCCTTGA 360
OY 965 CTGGGTGATTCACAGCTGACAGCAGGCAAGATCGTCTGACCCAGGAGAGAGAGCTG 1024
|||||
Db 361 CTGGGTGATTCACAGCTGACAGCAGGCAAGATCGTCTGACCCAGGAGAGAGAGAGCTG 420
OY 1025 GTGCAGCCCTTCAGCTCGCTGTTCGCCGAAGGTGAGTACATCGCAGGGCGGGTGGAC 1084
|||||
Db 421 GTGCAGCCCTTCAGCTCGCTGTTCGCCGAAGGTGAGTACATCGCAGGGCGGGTGGAC 480
OY 1085 CGGATGAGCAATATACGCTGTGGGCAATGTCTTGACACCGGCCCCACAGTGTCTCAAGCTC 1144
|||||
Db 481 CGGATGAGCAATATACGCTGTGGGCAATGTCTTGACACCGGCCCCACAGTGTCTCAAGCTC 540
OY 1145 GTGCTCTCTCCCGGCGGCTGTTCATCCGAGCAGAGAAATGAGAGAGCGGCTAGGC 1204
|||||
Db 541 GTGCTCTCTCCCGGCGGCTGTTCATCCGAGCAGAGAAATGAGAGAGCGGCTAGGC 600
OY 1205 TCTGCGAAGCTGTCCCGAGAAATGTCAAGCCGTATGTGTGATCAGAGAGTCCACCAAC 1264
|||||
Db 601 TCTGCGAAGCTGTCCCGAGAAATGTCAAGCCGTATGTGTGATCAGAGAGTCCACCAAC 660
OY 1265 GTGTGATCAATGTCATGACATCTTCTATCCCTTCCCAATCAAGAGAGAGAGAGAG 1324
|||||
Db 661 GTGTGATCAATGTCATGACATCTTCTATCCCTTCCCAATCAAGAGAGAGAGAGAGAG 720
OY 1325 CTCTGCTTCTCCCGGCAATGAATGACAGCGGCTTCTGCAATTTGTACAAATGACAC 1384
|||||
Db 721 CTCTGCTTCTCTCCCGGCAATGAATGACAGCGGCTTCTGCAATTTGTACAAATGACAC 780
OY 1385 G-CGCTTTTAAATCCCAAGG 1404
|||||
Db 781 GCCCGTTTAAATCCCAAGG 801
RESULT 9
BO689588 872 bp mRNA linear EST 15-JUL-2002
LOCUS BO689588
DEFINITION AGENCOURT_8345928 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6250957

5', mRNA sequence.
BO689588
VERSION BO689588.1 GI:21814904
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 872)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L1CM2394 row: n column: 14
High quality sequence stop: 654.
Location/Qualifiers
1..872
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6250957"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 192 a 258 c 247 g 174 t 1 others
ORIGIN

Query Match 29.9%; Score 783.2; DB 14; Length 872;
Best Local Similarity 99.4%; Pred. No. 7.5e-161;
Matches 796; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

OY 605 GCCTCTCTCTCTCATATTAACAGCAGCTGTGTGGCCCAACATCGAGACAGCGGAG 664
|||||
Db 1 GCCTCTCTCTCTCATATTAACAGCAGCTGTGTGGCCCAACATCGAGACAGCGGAG 60
OY 665 GAGCGGGGCTGACCTTCTGCGACCAAGTTTATTCATATGCTTGATGACCCCAAGTCT 724
|||||
Db 61 GAGCGGGGCTGACCTTCTGCGACCAAGTTTATTCATATGCTTGATGACCCCAAGTCT 120
OY 725 GCGGGTGTGGCCACTTGTCTATACAGAAAGTTCGACCGCTTACTGGGTACTGTGG 784
|||||
Db 121 GCGGGTGTGGCCACTTGTCTATACAGAAAGTTCGACCGCTTACTGGGTACTGTGG 180
OY 785 TGCCCCACAGCCTCTCTGGGAAGTTCAGAGGGCTCAAGACGCTCGAATCCTGTATGAG 844
|||||
Db 181 TGCCCCACAGCCTCTCTGGGAAGTTCAGAGGGCTCAAGACGCTCGAATCCTGTATGAG 240
OY 845 GAAGTCGATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAGAACAAAGG 904
|||||
Db 241 GAAGTCGATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAGAACAAAGG 300
OY 905 AAGAGGAGCTGTATCGGTATACCCCAAGGACAGGACAGCAAGATCCCAAGTTGCTTGA 964
|||||
Db 301 AAGAGGAGCTGTATCGGTATACCCCAAGGACAGGACAGCAAGATCCCAAGTTGCTTGA 360
OY 965 CTGGGTGATTCACAGCTGACAGCAGGCAAGATCGTCTGACCCAGGAGAGAGAGAGCTG 1024
|||||
Db 361 CTGGGTGATTCACAGCTGACAGCAGGCAAGATCGTCTGACCCAGGAGAGAGAGAGAGCTG 420

QY 1025 GTGAGCCCTTACGCTGCTGTTCCCGAAGGTGAGTACATGCCAGGCCGCGGTGACC 1084
|||||
Db 421 GTGAGCCCTTACGCTGCTGTTCCCGAAGGTGAGTACATGCCAGGCCGCGGTGACC 480
QY 1085 CGGATGGCAATATCGCTGGGCGATGTTCTGACAGCGGCCCGCAGCACTGCTCAGTCC 1144
|||||
Db 481 CGGATGGCAATATCGCTGGGCGATGTTCTGACAGCGGCCCGCAGCACTGCTCAGTCC 540
QY 1145 GTCTCTCTCCCGCGCGCTTTCATCCCGACAGAGTAAGAGAGCAGCGCTAGCC 1204
|||||
Db 541 GTCTCTCTCCCGCGCGCTTTCATCCCGACAGAGTAAGAGAGCAGCGCTAGCC 600
QY 1205 TCTGCGAGAGCTGTCCCGAGAAATGTCAGCGCATGTGTGTGATGAGAGTCAAC 1264
|||||
Db 601 TCTGCGAGAGCTGTCCCGAGAAATGTCAGCGCATGTGTGTGATGAGAGTCAAC 660
QY 1265 GTCTGATCAATGTTCAATGACATCTTCTATCCCTTCCCGCATAGAGAGAGAGAG 1324
|||||
Db 661 GTCTGATCAATGTTCAATGACATCTTCTATCCCTTCCCGCATAGAGAGAGAGAG 720
QY 1325 CTCTGCTTCTCCGCGCAATGAATGCAAGA -CGGCTTCTGCGCATTTGTCAAAAGTCA 1383
|||||
Db 721 CTCTGCTTCTCCGCGCAATGAATGCAAGACCGGCTTCTGCGCATTTGTCAAAAGTCA 780
QY 1384 CGCGGTTTTAAATCCAGG 1404
|||||
Db 781 CGCGGTTTTAAATCCAGG 801

RESULT 10
BM461814 866 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6418408 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5533867
DEFINITION 5', mRNA sequence.
ACCESSION BM461814
VERSION BM461814.1 GI:18510854
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgrabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
DNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM1219 row: g column: 20
High quality sequence stop: 627.
Location/Qualifiers
1..866
source

FEATURES
source
1..866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5533867"
/clone_id="NIH_MGC_71"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: Nct1;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
188 a 269 c 245 g 164 t

BASE COUNT 188 a 269 c 245 g 164 t
ORIGIN
Query Match 29.88; Score 779.8; DB 13; Length 866;
Best Local Similarity 98.86; Pred. No. 4.2e-160;

Matches 817; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 128 GCGAGCCGCAAGTACTGGGCTCATTTGTCAACAAGCGGCCCGCAGACTTCCAGTTGTG 187
|||||
Db 11 GCGAGCCGCAAGTACTGGGCTCATTTGTCAACAAGCGGCCCGCAGACTTCCAGTTGTG 70
QY 188 CAGAAGCGGATGAGTCTGGGCCCTCCACCGCTTACTAGTACGTGGGAATGCCATAT 247
|||||
Db 71 CAGAAGCGGATGAGTCTGGGCCCTCCACCGCTTACTAGTACGTGGGAATGCCATAT 130
QY 248 GCGAGCCGCAAGTACTGGGCTCATTTGTCAACAAGCGGCCCGCAGACTTCCAGTTGTG 307
|||||
Db 131 GCGAGCCGCAAGTACTGGGCTCATTTGTCAACAAGCGGCCCGCAGACTTCCAGTTGTG 190
QY 308 CTGCTGCTCTGCTCTGGAAGCAGATGCTGGATATTTCCAGGCGCAGCCACCATGAG 367
|||||
Db 191 CTGCTGCTCTGCTCTGGAAGCAGATGCTGGATATTTCCAGGCGCAGCCACCATGAG 250
QY 368 GTCTACTCTGGGAGGAGAGCTGTGAGGAGCGGAAGCGCTGGGGTCTTGGGATC 427
|||||
Db 251 GTCTACTCTGGGAGGAGAGCTGTGAGGAGCGGAAGCGCTGGGGTCTTGGGATC 310
QY 428 ACCTCTACGACTTCCACAGGAGAGTGGCTCTTCTCTTCCAGGCGCAGCAAGCTC 487
|||||
Db 311 ACCTCTACGACTTCCACAGGAGAGTGGCTCTTCTCTTCCAGGCGCAGCAAGCTC 370
QY 488 TTCCACTGTGCGAGCGCGGCAAGAGCGCTTCTGTTGTCCTTATGAAACCGCTGAA 547
|||||
Db 371 TTCCACTGTGCGAGCGCGGCAAGAGCGCTTCTGTTGTCCTTATGAAACCGCTGAA 430
QY 548 ATCAAGACCCAGTCTGAGGCGCCCGGATGAGACCCCAAAATCTGCTCCGCGACCTGCG 607
|||||
Db 431 ATCAAGACCCAGTCTGAGGCGCCCGGATGAGACCCCAAAATCTGCTCCGCGACCTGCG 490
QY 608 TTCTTCTCTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 667
|||||
Db 491 TTCTTCTCTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 550
QY 668 CGGCGGCTGACCTTCTGCAACCAAGTTTATCCAAATGCTGATGACCCCAAGTGTGCG 727
|||||
Db 551 CGGCGGCTGACCTTCTGCAACCAAGTTTATCCAAATGCTGATGACCCCAAGTGTGCG 610
QY 728 GGTGTGCGCACCTTCTGCAACCAAGTTTATCCAAATGCTGATGACCCCAAGTGTGCG 787
|||||
Db 611 GGTGTGCGCACCTTCTGCAACCAAGTTTATCCAAATGCTGATGACCCCAAGTGTGCG 670
QY 788 CCCACAGCTCTCTGGAAGTTGAGAGGCTTCAAGAGCGTGGCAATCTGTATGAGAA 847
|||||
Db 671 CCCACAGCTCTCTGGAAGTTGAGAGGCTTCAAGAGCGTGGCAATCTGTATGAGAA 730
QY 848 GTGATGAGT -CCGAGGTGAGGTCATTTCAAGTCCCTCTCTGCTAGAG -AAAGA 905
|||||
Db 731 GTGATGAGTCCGAGGTGAGGTCATTTCAAGTCCCTCTCTGCTAGAG -AAAGA 790
QY 906 AGACGAGCTGATGCTA -CCCCAGACAGCAGCAAGATCCCA 951
|||||
Db 791 AGACGAGCTGATGCTA -CCCCAGACAGCAGCAAGATCCCA 837

RESULT 11
BQ949519 910 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8794758 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:5374913
DEFINITION 5', mRNA sequence.
ACCESSION BQ949519
VERSION BQ949519.1 GI:22364997
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 910)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT Robert Strausberg, Ph.D.
COMMENT Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
DNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LCM2554 row: k column: 10
High quality sequence stop: 668.
Location/Qualifiers
1. 910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6374913"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 190 a 294 c 261 g 165 t
ORIGIN

Query Match 29.4%; Score 770.6; DB 14; Length 910;
Best Local Similarity 97.8%; Pred. No. 4.4e-158;
Matches 834; Conservative 0; Mismatches 14; Indels 5; Gaps 5;

11 ATGGCACCACCGGAGACCCCAACGCGGAGCGAGCGAGCGCCACAGATGACCGG 70
|||||
37 ATGGCACCACCGGAGACCCCAACGCGGAGCGAGCGCGCCACAGATGACCGG 96
|||||
71 GCCGCCGCTTCAGGTGAGAGCAGTCTGGGAGCGGCGTCCGAGACATCCACGCG 130
|||||
97 GCCGCCGCTTCAGGTGAGAGCAGTCTGGGAGCGGCGTCCGAGACATCCACGCG 156
|||||
131 AGCCGCAAGTACTCGGCGCTATTGTCAACAAGCGCGCCACAGATTCAGTTTGTGAC 190
|||||
157 AGCCGCAAGTACTCGGCGCTATTGTCAACAAGCGCGCCACAGATTCAGTTTGTGAC 216
|||||
191 AAGACGATGAGTGGGCGCCACTCCACCGCTCTACTACTGGAATGCGATATGCG 250
|||||
217 AAGACGATGAGTGGGCGCCACTCCACCGCTCTACTACTGGAATGCGATATGCG 276
|||||
251 AGCCGAGAACTCCCTCTACTCTGAGATTCCCAAGAGTCCGGAAGAGGCTCG 310
|||||
277 AGCCGAGAACTCCCTCTACTCTGAGATTCCCAAGAGTCCGGAAGAGGCTCG 336
|||||
311 CTGCTCTCTCTCTGAGACAGATGCTGATCATTTTCAGAGCCACGCCCATATGGGTC 370
|||||
337 CTGCTCTCTCTCTGAGACAGATGCTGATCATTTTCAGAGCCACGCCCATATGGGTC 396
|||||
371 TACTCTCGGAGAGAGTGTGAGGGAGGGAAGCGCTGGGGGTCTTGGCATACCC 430
|||||
397 TACTCTCGGAGAGAGTGTGAGGGAGGGAAGCGCTGGGGGTCTTGGCATACCC 456
|||||
431 TCCTACGACTTCACAGAGAGAGTGGCTCTTCTCTTCCAGAGCCAGCAACGCTCTTC 490
|||||
457 TCCTACGACTTCACAGAGAGAGTGGCTCTTCTCTTCCAGAGCCAGCAACGCTCTTC 516
|||||
491 CACTCTCGGAGAGCGGCGCAAGAGGCTTCATGTTGTCCTCATGAACCGCTGGAATC 550
|||||
517 CACTCTCGGAGAGCGGCGCAAGAGGCTTCATGTTGTCCTCATGAACCGCTGGAATC 576
|||||
551 AAGACCAAGTGTCAAGGGCGCGGATGGAACCCCAAAATCTGCTTGGCAGACCTGCTTC 610
|||||

|||||
Db 577 AAGACCAAGTGTCAAGGGCGCGGATGGAACCCCAAAATCTGCTTGGCAGACCTGCTTC 636
|||||
Qy 611 TTCTCTCTCATCAATACAGAGAGCTGTGGTGGCCAACTCGACAGAGAGAGCGG 670
|||||
Db 637 TTCTCTCTCATCAATACAGAGAGCTGTGGTGGCCAACTCGACAGAGAGAGCGG 696
|||||
Qy 671 CGGCTGACCTTCTGCGACCAAGGTTTATCCAAATGTCTGGATGACCCCAAGTCTGC 729
|||||
Db 697 CGGCTGACCTTCTGCGCA-CAGGTTTATCCAAATGTCTGGATGACCCCAAGTCTGC 755
|||||
Qy 730 TGTGGCCACTTCTGTCATACAGAGAGAGTTCGA-CCGCTTCACTGGTACTGTGTGCC 788
|||||
Db 756 TGTGGGCACTTCTGTCATACAGAGAGAGTTCGA-CCGCTTCACTGGTACTGTGTGCC 815
|||||
Qy 769 CCACAG-CCTCTCTGGGAAGGTTTCAGAGGCGCTCAAGAGCTG-CGAATCTGTATGAGA 846
|||||
Db 816 CCACAGCCTCTCTGGGAGGTTTCAGAGGCGCTCAAGAGCTGCGCAATCTGTATGAGA 875
|||||
Qy 847 AGTCATGAGTCC 859
|||||
Db 876 AGTCATGAGTCC 888
|||||

RESULT 12
B0642814 902 bp mRNA linear EST 15-JUL-2002
LOCUS B0642814
DEFINITION AGENCOURT_8285906 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:6292546
5', mRNA sequence.
ACCESSION B0642814
VERSION B0642814.1 GI:21766986
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 902)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Lou Strand
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LCM2494 row: c column: 11
High quality sequence stop: 637.
Location/Qualifiers
1. 902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6292546"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 202 a 269 c 255 g 175 t 1 others
ORIGIN

Query Match 29.3%; Score 767.8; DB 14; Length 902;
Best Local Similarity 95.9%; Pred. No. 1.8e-157;

Matches	809: Conservative	0: Mismatches	33: Indels	2: Gaps
QY	852	ATGAGTCGAGAGTGGAGGATCATTCACGTCCTCCCTCTCTCCGCTAGACAAAGAGACGG	911	
Db	1	ATGAGTCGAGAGTGGAGGATCATTCACGTCCTCCCTCTCTCCGCTAGACAAAGAGACGG	60	
QY	912	ACTCGATGGTACCCACAGACAGGACGACAAAGATCCCAAGATTGGCTTGAAGTGGCTG	971	
Db	61	ACTCGATGGTACCCACAGACAGGACGACAAAGATCCCAAGATTGGCTTGAAGTGGCTG	120	
QY	972	AGTTCCAGACTGACAGCCAGGGCCAGAGATGCTTCGACCCAGGAGAAAGAGTGTGCACG	1031	
Db	121	AGTTCCAGACTGACAGCCAGGGCCAGAGATGCTTCGACCCAGGAGAAAGAGTGTGCACG	180	
QY	1032	CCTTCAGCTGCTGTTCCTCGAAGTGTGATACATGCCAGGGCCGGGTGGACCCGGGATG	1091	
Db	181	CCTTCAGCTGCTGTTCCTCGAAGTGTGATACATGCCAGGGCCGGGTGGACCCGGGATG	240	
QY	1092	GCAAAATAGCCCTGGGGGCATGTCCTGGACAGGGCCCGACAGATGGCTCCAGCTGCTCC	1151	
Db	241	GCAAAATAGCCCTGGGGGCATGTCCTGGACAGGGCCCGACAGATGGCTCCAGCTGCTCC	300	
QY	1152	TCCCCCGGCGCCTGTTCATCCCGACACAGAAATGAGAGAGCAGCGGCTTGCCA	1211	
Db	301	TCCCCCGGCGCCTGTTCATCCCGACACAGAAATGAGAGAGCAGCGGCTTGCCA	360	
QY	1212	GAGCTGTCCCAAGAAATGTCAGCCGATATGTGTACAGAGAGGTACACCAACGTCGTGA	1271	
Db	361	GAGCTGTCCCAAGAAATGTCAGCCGATATGTGTACAGAGAGGTACACCAACGTCGTGA	420	
QY	1272	TCAATGTCATACAACTCTATCCCTCCCTCCCAATCAGAGGAGAGACGAGCTCTGCT	1331	
Db	421	TCAATGTCATACAACTCTATCCCTCCCTCCCAATCAGAGGAGAGACGAGCTCTGCT	480	
QY	1332	TTCCTCGCGCCCAATGAATGCAAGACCGGCTCTGCCATTTGTAAAGTCACACGCGCTT	1391	
Db	481	TTCCTCGCGCCCAATGAATGCAAGACCGGCTCTGCCATTTGTAAAGTCACACGCGCTT	540	
QY	1392	TAAAAATCCCAAGGCTACAGTGTGAGTGAAGCCCTTCAGCCCCGGGGAGATGATTTAACT	1451	
Db	541	TAAAAATCCCAAGGCTACAGTGTGAGTGAAGCCCTTCAGCCCCGGGGAGATGATTTAACT	600	
QY	1452	GCCCATTAAGGAAGAGATTGCTGACCCAGGCGGTGAATGGGAGGATTTGGCGAGGCAC	1510	
Db	601	GCCCATTAAGGAAGAGATTGCTGACCCAGGCGGTGAATGGGAGGATTTGGCGAGGCAC	660	
QY	1511	GGCCTCAAGATTTGGGTCAATGAGAGAGACAAGCTGTACTTCCAGGGACCAAGAAGAC	1570	
Db	661	GGCCTCAAGATTTGGGTCAATGAGAGAGACAAGCTGTACTTCCAGGGACCAAGAAGAC	720	
QY	1571	ACGCGCGTGGAGACACACCTCTACGTGTGATGAGAGCGCGCGGAGATCGTACGC	1630	
Db	721	ACGCGCGTGGAGACACACCTCTACGTGTGATGAGAGCGCGCGGAGATCGTACGC	780	
QY	1631	CTCACCAGGCGCGGCTTC-TCCCATAGCTGCTCCATGAGCCAGAACTTGACATGTTGCT	1689	
Db	781	CTCACCAGCTCCGGGTTCTTCCCTAAAGCTGTCATGAAGCCAAAAATTTCCACATGTTT	840	
QY	1690	CAGC 1693		
Db	841	CGGC 844		
RESULT 13				
BM915935				
LOCUS	BM915935	1051 bp	mRNA	linear
DEFINITION	AGENCOURT_6639540 NIH_MGC_41 Homo sapiens	CDNA clone	IMAGE:5482205	
ACCESSION	BM915935			
VERSION	BM915935.1	GI:19366314		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 1051)	Enkaryota: Metazoa; Chordata; Craniata; Vertebrata: Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
1	NH-MGC	http://mgc.nci.nih.gov/.		
1	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)		
1	Contact: Robert Strausberg, Ph.D.	Email: cgapbs-remail.nih.gov		
1	Tissue Procurement: DCTD/DRP	CDNA Library Preparation: Rubin Laboratory		
1	DNA Sequencing by: Agencourt Bioscience Corporation	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
1	Plate: L12CM2007	row: o column: 06		
1	High quality sequence status: 677.	Location/Qualifiers		
1	1. 1051	/organism="Homo sapiens"		
1	/db_xref="taxon:9606"	/clone="IMAGE:5482205"		
1	/clone_lib="NH_MGC_41"	/tissue_type="amelanotic melanoma, cell line"		
1	/lab_host="DH10B (phage-resistant)"	/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally clone into EcoRI/XhoI sites using the following 5' adaptor: GGCCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NH_MGC Library."		
BASE COUNT	233 a	329 c	300 g	189 t
ORIGIN				
Query Match	29.3%	Score 767.4;	DB 14;	Length 1051;
Best Local Similarity	96.0%;	Ident. No. 2,3e15;		
Matches 853;	Conservative 0;	Mismatch 26;	Indels 10;	Gaps 6;
QY 349	GGCCAGCCCCACATGAGGGGTCTACTCTCGGAGAGAGAGCTGCTGAGGAGCGGAACG	408		
Db 1	GGCCAGCCCCACATGAGGGGTCTACTCTCGGAGAGAGAGCTGCTGAGGAGCGGAACG	60		
QY 409	CTGTGGGGTCTTGGGCATCACCTCTCTACGACTTCCACAGCGAGTGGCTTCTCTT	468		
Db 61	CTGTGGGGTCTTGGGCATCACCTCTCTACGACTTCCACAGCGAGTGGCTTCTCTT	120		
QY 469	CCAGGCGACGAACAGCCTCTTCCACTGTGCGAGGCGGCAAGAACGCTTATGCTGTC	528		
Db 121	CCAGGCGACGAACAGCCTCTTCCACTGTGCGAGGCGGCAAGAACGCTTATGCTGTC	180		
QY 529	CCCTATGAACCGGTGGAATCAACACCAAGTGCAGAGGCGCCGGATGGAGCCCAAAAT	588		
Db 181	CCCTATGAACCGGTGGAATCAACACCAAGTGCAGAGGCGCCGGATGGAGCCCAAAAT	240		
QY 589	CTGGCTTCCGACACCTGCTCTTCTTCCCTTCATCAATAACAGCAGCTTGGGTGGCAA	648		
Db 241	CTGGCTTCCGACACCTGCTCTTCTTCCCTTCATCAATAACAGCAGCTTGGGTGGCAA	300		
QY 649	CATGAGACAGGCGAGAGAGCGGCGCTGACCTTCTGCAACCAAGGTTTATCCAAATGCTT	708		
Db 301	CATGAGACAGGCGAGAGAGCGGCGCTGACCTTCTGCAACCAAGGTTTATCCAAATGCTT	360		
QY 709	GGATGACCCCAAGTCTGGGGGTGTGGCCACCTTCTCATACAGGAAGATTGGACCGCTT	768		
Db 361	GGATGACCCCAAGTCTGGGGGTGTGGCCACCTTCTCATACAGGAAGATTGGACCGCTT	420		
QY 769	CACGTGGTACGTGGTGTGTCGCCACAGCCTTCCGGGAAGTTAGAGGGGCTCAAGACGT	828		
Db 421	CACGTGGTACGTGGTGTGTCGCCACAGCCTTCCGGGAAGTTAGAGGGGCTCAAGACGT	480		
QY 829	GCGAATCTCTATGAGGAAGTCGATGAGTCCGAGGTGAGAGGTCAATTCAGCTCCCTTCC	888		

|||||
Db 481 GCGAATCCTGATGAGGAAGTCGATGATCCGAGGTGAGAGTCATTCACGTCCCTCTCC 540
Qy 889 TGCCTAGAAAAGAAAGAGAGGACTGTATCGTATACCCCGAGAGCAGCAGAAATCC 948
Db 541 TGCCCTAGAAAAGAAAGAGAGGACTGTATCGTATACCCCGAGAGCAGCAGAAATCC 600
Qy 949 CAGATTCCCTTGAACAGCGCTGAGTTCCAGACTGACAGCAGGCGCAAGATCGTCTGAC 1008
Db 601 CAGATTTCCTTGAACAGCGCTGAGTTCCAGACTGACAGCAGGCGCAAGATCGTCTGAC 660
Qy 1009 CCAGAGAGAGAGCTGTGTGACAGCCCTTCAGCTGCTGTTCGCCGAAGTGTACATACATGC 1068
Db 661 CCAGAGAGAGAGCTGTGTGACAGCCCTTCAGCTGCTGTTCGCCGAAGTGTACATACATGC 720
Qy 1069 CAGGCGCCGGTGG--ACCCGGGATGGCAATPAG-CTTGCGCATGTTCTTCGACCGG- 1123
Db 721 GGGGGGCGCGGTGGAAACCCCGGAATGGCAATAGCCCTGGGGCCATGTTCTTCGACCGGC 780
Qy 1124 CCCAGCAGTGGCTCCAGCTGTCTCTCTCCGCC-GGCCCTGTTCAATCCGAGAC--A 1180
Db 781 CCCAGCAGTGGCTCCAGCTGTCTCTCTCCGCCCGCGGCGCTGTTCAATCCGAGACCCCA 840
Qy 1181 GAGATGAGAGAGCAGCGGCT--AGCCTTCGCCAGAGCTGTCCCGCAGAA 1227
Db 841 AGAATAGAGAGCAGCGGCTTAACCTCTGCGCAGAACTGGCCCAAGAA 889

RESULT 14
LOCUS B1223668 787 bp mRNA linear EST 11-JUL-2001
DEFINITION 60294358.F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5106357 5',
mRNA sequence.
ACCESSION B1223668
VERSION B1223668.1 GI:14677112
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 787)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNLT Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L14M11257 row: b column: 22
High quality sequence stop: 787.

FEATURES

location/Qualifiers
1..787

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5106357"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 164 a 263 c 220 g 140 t
ORIGIN

Query Match 29.0%; Score 757.8; DB 13; Length 787;
Best local similarity 99.0%; Pred. No. 2,6e-155;
Matches 773; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Qy 11 ATGGCCACACCGGGACCCCAACGGCCGACGAGGCGAGCAGCCGCCACAGATGACCG 70
Db 7 ATGGCCACACCGGGACCCCAACGGCCGACGAGGCGAGCAGCCGCCACAGATGACCG 66
Qy 71 GCCCCCGCTTCCAGGTGACAGACACTGCTGGGACGGGCTCCGAGCATATCCACGGC 130
Db 67 GCCCGCCCTTCCAGGTGACAGACACTGCTGGGACGGGCTCCGAGCATATCCACGGC 126
Qy 131 AGCCGCAAGTACTGGGCTCATTTGTCAACAAGGCGCCCAAGCACTTCCAGTTGTGAG 190
Db 127 AGCCGCAAGTACTGGGCTCATTTGTCAACAAGGCGCCCAAGCACTTCCAGTTGTGAG 186
Qy 191 AAGAGGATGAGTCTGGGCCCCCATCCACCGGCTCTACTACTCTGGGAATGCCATATGCG 250
Db 187 AAGAGGATGAGTCTGGGCCCCCATCCACCGGCTCTACTACTCTGGGAATGCCATATGCG 246
Qy 251 AGCCGAGAACTCCCTCTACTCTGAGATTCACAAGAGGTCGGAAAGAGAGCTCTG 310
Db 247 AGCCGAGAACTCCCTCTACTCTGAGATTCACAAGAGGTCGGAAAGAGAGCTCTG 306
Qy 311 CTGCTCTGTCTGTGGAAGCAGATCTGATCATTTCCAGGCCACGCCACCATGGGGTC 370
Db 307 CTGCTCTGTCTGTGGAAGCAGATCTGATCATTTCCAGGCCACGCCACCATGGGGTC 366
Qy 371 TACTCTGGGAGAGAGAGACTCTGAGGAGCGGAAAGCCCTGGGGGTCTTGGCATCAC 430
Db 367 TACTCTGGGAGAGAGACTCTGAGGAGCGGAAAGCCCTGGGGGTCTTGGCATCAC 426
Qy 431 TCCTACGACTTCCACAGGAGAGTGGCTTTCCTCTCCAGGCCAGCAACAGCTCTTC 490
Db 427 TCCTACGACTTCCACAGGAGAGTGGCTTTCCTCTCCAGGCCAGCAACAGCTCTTC 486
Qy 491 CACTGTGCGAGCGGGCGGCAAGACGGCTTCATGTGTCCCTATGAAACCGCTGGAATTC 550
Db 487 CACTGTGCGAGCGGGCGGCAAGACGGCTTCATGTGTCCCTATGAAACCGCTGGAATTC 546
Qy 551 AAGACCCAGTCTCAGGGCCCCCGATGAGACCCCAAAATCTGCTCCGCGACCTGCTTC 610
Db 547 AAGACCCAGTCTCAGGGCCCCCGATGAGACCCCAAAATCTGCTCCGCGACCTGCTTC 606
Qy 611 TTCTCTTCATCAATAAAGAGGACTGTGGTGGGCAACATCGAGACAGCGAGAGCGG 670
Db 607 TTCTCTTCATCAATAAAGAGGACTGTGGTGGGCAACATCGAGACAGCGAGAGCGG 666
Qy 671 CGCTGACCTTCTGCGACCAAGGTTATCAATGTCTGATGACCCCAAGTCTGGGGT 730
Db 667 CGCTGACCTTCTGCGACCAAGG--TTATCAATGTCTGATGACCCCAAGTCTGGGGT 725
Qy 731 GTGGCCACCTTCGTCATACAGGAAGTTGACACCGCTTCACCTGAGTGTGTCGCC 790
Db 726 GTGGCCACCTTCGTCATACAGGAAGTTGACACCGCTTCACCTGAGTGTGTCGCC 785
Qy 791 A 791
Db 786 A 786

RESULT 15
LOCUS B0685669 938 bp mRNA linear EST 15-JUL-2002
DEFINITION AGNCOURT_8176462 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6252323
5', mRNA sequence.
ACCESSION B0685669
VERSION B0685669.1 GI:21810985
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 938)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9abs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>

Plate: LICM2398 row: 9 column: 12
High quality sequence stop: 646.

FEATURES
source Location/Qualifiers

1. 938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6252323"
/clone_id="NIH_MGC_110"
/issue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 209 a 278 c 262 g 189 t

ORIGIN

Query Match 28.7%; Score 750; DB 14; Length 938;

Best Local Similarity 97.9%; Pred. No. 1.4e-153;

Matches 781; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

```
QY 605 GCCTTCTTCCTTCATCAATACAGGACCTGTGGTGCCCAACATGACAGACGGCAG 664
      |||||||
Db 1 GCCTTCTTCCTTCATCAATACAGGACCTGTGGTGCCCAACATGACAGACGGCAG 60

QY 665 GAGGGGGGGGACCTCTGCGACCAAGTTTATCCAAATGCTCGATGCCCAAGTCT 724
      |||||||
Db 61 GAGGGGGGGGACCTCTGCGACCAAGTTTATCCAAATGCTCGATGCCCAAGTCT 120

QY 725 GCGGGTGTGGCACCTTCGTATACAGGAAGTTTCAGCCGCTTCACTGGGTACTGGTG 784
      |||||||
Db 121 GCGGGTGTGGCACCTTCGTATACAGGAAGTTTCAGCCGCTTCACTGGGTACTGGTG 180

QY 785 TGGCCCAACAGCCTCTGGAAGTTTCAGAGGGCCTCAAGACGCTGCGAATCCTGTATGAG 844
      |||||||
Db 181 TGGCCCAACAGCCTCTGGAAGTTTCAGAGGGCCTCAAGACGCTGCGAATCCTGTATGAG 240

QY 845 GAAGTCATGATCGAGGTGGAGTGATTCACGTCCCTCTGCGCTAGAGAAAGG 904
      |||||||
Db 241 GAAGTCATGATCGAGGTGGAGTGATTCACGTCCCTCTGCGCTAGAGAAAGG 300

QY 905 AAGACGACTCGTATCGGTACCCAGAGACAGCAAGAAATCCCAAGATTGCTTGA 964
      |||||||
Db 301 AAGACGACTCGTATCGGTACCCAGAGACAGCAAGAAATCCCAAGATTGCTTGA 360

QY 965 CTGGCTAGTTCCAGACTGACAGGCAAGATGCTCTGACCCAGAGAGAGAGCTG 1024
      |||||||
Db 361 CTGGCTAGTTCCAGACTGACAGGCAAGATGCTCTGACCCAGAGAGAGAGCTG 420

QY 1025 GTGAGCCCTTCAGCTGGCTGTTCGGAAGTGGATACATCCGCCA -GGGCCGGGTGAC 1083
      |||||||
Db 421 GTGAGCCCTTCAGCTGGCTGTTCGGAAGTGGATACATCCGCCAAGGGGTGAC 480

QY 1084 CCGGATGCAAAATACCTGAGGCAATGTTCTTGACCGGCCAGAGAGTGGCTCCAGCT 1143
      |||||||
Db 481 CCGGATGCAAAATACCTGAGGCAATGTTCTTGACCGGCCAGAGAGTGGCTCCAGCT 540

QY 1144 CGTCTCTCTCCCGGCCCTGTTCATCCGAGCAGAGAAATGAGGACGCGGTAGC 1203
      |||||||
```

```
Db 541 CGTCTCTCTCCCGGCCCTGTTCATCCGAGCAGAGAAATGAGGACGCGGTAGC 600
QY 1204 CTGAGCAGAGCTGTCCCAAGAAATGTCAGCCGTATGTGTGATGAGAGAGTCAACCA 1263
      |||||||
Db 601 CTGAGCAGAGCTGTCCCAAGAAATGTCAGCCGTATGTGTGATGAGAGAGTCAACCA 660

QY 1264 CGTCTGATCAATGTTTCATGACATCTTCTATCCCTTCCCAATCAGA -GGAGAGAGC 1322
      |||||||
Db 661 CGTCTGATCAATGTTTCATGACATCTTCTATCCCTTCCCAATCAAAAGGAGAGAGC 720

QY 1323 AGCTCTGCTTTCCTCCGGCCCAATGCAAGACCGGCTTGTGCATTTGTACAAACTCA 1382
      |||||||
Db 721 AGCTCTGCTTTCCTCCGGCCCAATGCAAGACCGGCTTGTGCATTTGTACAAACTCA 780

QY 1383 CCGGCCCTTAAATCCC 1400
      || ||||| ||
Db 781 CCGGCCCTTAAATATCC 798
```

Search completed: December 12, 2002, 10:36:46
Job time : 2402 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 08:32:37 : Search time 74 Seconds
(without alignments)
10845.582 Million cell updates/sec

Title: US-09-976-674-4

Sequence: 1 caagcttaccatgcccaccca.....tctgagcggccgcgataccg 2617

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2.6/prodata1/ina/5A.COMB.seq:*
2: /cgn2.6/prodata1/ina/5B.COMB.seq:*
3: /cgn2.6/prodata1/ina/6A.COMB.seq:*
4: /cgn2.6/prodata1/ina/6B.COMB.seq:*
5: /cgn2.6/prodata1/ina/PCTUS.COMB.seq:*
6: /cgn2.6/prodata1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152.8	5.8	823	4	US-09-280-116-171
2	127	4.9	612	4	US-09-392-184-31
3	87.6	3.3	5496	4	US-09-462-284-1
4	68.2	2.6	543	4	US-09-221-017B-253
5	55.6	2.1	2924	5	PCT-US93-07923-1
6	46.6	1.8	2815	1	US-08-230-91A-1
7	46.6	1.8	2815	1	US-08-619-280A-1
8	46.6	1.8	2815	2	US-08-940-391-1
9	45.4	1.7	1575	3	US-08-858-876A-1
10	45.4	1.7	1575	4	US-09-472-880-1
11	45.4	1.7	8438	1	US-07-945-283-1
12	45	1.7	1342	3	US-08-832-399-1
13	45	1.7	1342	4	US-09-372-498-1
14	44.4	1.7	657	4	US-09-221-017B-646
15	44.4	1.7	3624	1	US-07-951-715A-6
16	44.4	1.7	3624	2	US-08-459-448A-6
17	44.4	1.7	3624	3	US-08-459-595A-6
18	44.4	1.7	3624	3	US-08-459-504B-6
19	44.4	1.7	3624	3	US-08-459-444-6
20	44.4	1.7	3624	3	US-09-053-549-7
21	44.4	1.7	3624	4	US-09-547-422-6
22	44.4	1.7	8854	3	US-09-053-549-1
23	43.8	1.7	2214	3	US-08-864-038A-1
24	43.8	1.7	3331	3	US-08-864-038A-2
25	43.8	1.7	3331	3	US-08-864-038A-4
26	43.6	1.7	43280	2	US-08-804-227C-1
27	43.2	1.7	962	4	US-08-765-907A-16

28	43.2	1.7	1052	2	US-08-403-852D-10	Sequence 10, Appl
29	43.2	1.7	1052	3	US-08-510-646B-10	Sequence 10, Appl
30	43.2	1.7	1052	4	US-09-231-818-10	Sequence 10, Appl
31	42.2	1.6	1835	4	US-09-417-704-2	Sequence 2, Appl1
32	42	1.6	7898	4	US-08-984-709A-49	Sequence 49, Appl
33	42	1.6	49272	1	US-08-614-770A-1	Sequence 1, Appl1
34	41.8	1.6	2849	3	US-08-809-286B-1	Sequence 1, Appl1
35	41.6	1.6	926	4	US-09-500-569-19	Sequence 19, Appl
36	41.4	1.6	8051	2	US-08-576-626A-2	Sequence 2, Appl1
37	41.2	1.6	1829	1	US-08-035-558-2	Sequence 2, Appl1
38	41.2	1.6	1829	2	US-08-682-847-3	Sequence 3, Appl1
39	41	1.6	1302	4	US-09-372-422A-27	Sequence 3, Appl1
40	41	1.6	4960	4	US-09-907-843-3	Sequence 3, Appl1
41	40.8	1.6	1554	1	US-08-469-486-1	Sequence 1, Appl1
42	40.8	1.6	1554	2	US-08-469-658-1	Sequence 1, Appl1
43	40.8	1.6	2481	1	US-08-324-243-35	Sequence 35, Appl
44	40.8	1.6	2481	1	US-08-532-390-35	Sequence 35, Appl
45	40.8	1.6	2481	3	US-08-717-294-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-09-280-116-171
: Sequence 171, Application US/09280116A
: Patent No. 6331427
: GENERAL INFORMATION:
: APPLICANT: Robison, Keith E.
: TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
: FILE REFERENCE: 5800-24, 035800/176965
: CURRENT APPLICATION NUMBER: US/09/280, 116A
: CURRENT FILING DATE: 1999-03-26
: NUMBER OF SEQ ID NOS: 268
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 171
: LENGTH: 823
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: prolyl oligopeptidases
US-09-280-116-171

Query Match 5.8%; Score 152.8; DB 4; Length 823;
Best Local Similarity 62.5%; Pred. No. 3e-25;
Matches 255; Conservative 0; Mismatches 152; Indels 1; Gaps 1;

QY	2103	GCCAGGTGAGATGAGAGACGAGTGG-AGGCGCTGCAGTTCGTGGCCGAGAAATGAGGC	2161
DB	1	CTCAATATGAATATGACGATGAGTGGAAAGACATCCATATCTAGCTTCGATATGAT	60
QY	2162	TTTCATGACCTGAGCCGAGTTCATCCATGCGTGTCTTACGGGGCTTCCTGCTC	2221
DB	61	TTTCATGACTTAGAATGCGTGGCATCCAGCGGTGCTCTTATGAGATGATCTCCCTG	120
QY	2222	ATGGGGCTTAATCCACAGCCCGAGGTTCACAGTGGCCATCCGGGGTCCCGGCTGACC	2281
DB	121	ATGGCATTAATGACGAGGTGATATCTTCAGGGTGTCTATGCTGGGCTCCCACT	180
QY	2282	GTCGTGATGCGCTGACACAGGATGACACTGAGCGCTCATGAGAGTCCGTGAGAAC	2341
DB	181	CTGTGATCTCTTATGATACAGGATACAGGACCTTATATGGGTCCACCTGACCAAGAT	240
QY	2342	CAGCAGGCTATGAGCGGCTTCGTCGCCCTGACGACGTGAGAAAGTCCCAATGAGCC	2401
DB	241	GAAAGGCTTATCTTATGATGATGTCGTCGATGCAAGCAAGCAAGAAAGTCCCTGCAACCA	300
QY	2402	AACGCTGCTTATCTTACGAGGCTCTCTGACGAAACGTCGACTTTTCCACCAAC	2461
DB	301	AATGCTTACTGCTCTTACATGATGTTCTCTGATGAGATGTCATTTGACATACAGT	360
QY	2462	TTCTGCTGCCCACTGATCCGAGCAAGAACTTACAGCTTCAG	2509


```
? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94304-1018
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows
? SOFTWARE: FastSeq for Windows Version 2.0b
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/221,017B
? FILING DATE: 23-DEC-1998
? CLASSIFICATION:
? APPLICATION NUMBER: P1182
? FILING DATE: 31-DEC-1997
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: P1546
? FILING DATE: 30-JAN-1998
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: P2911
? FILING DATE: 09-APR-1998
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/AU98/01023
? FILING DATE: 10-DEC-1998
? ATTORNEY/AGENT INFORMATION:
? NAME: Monroy, Gladys H
? REGISTRATION NUMBER: 32,430
? REFERENCE/DOCKET NUMBER: 27340-20021.00
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650-813-5600
? TELEFAX: 650-494-0792
? TELEX: 706141
?
? INFORMATION FOR SEQ ID NO: 253:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 543 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: circular
? MOLECULE TYPE: DNA (genomic)
? HYPOTHETICAL: NO
? ANTI-SENSE: UNKNOWN
? ORIGINAL SOURCE:
? ORGANISM: PORPHYROMONAS GINGIVALIS
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1...543
?
? US-09-221-017B-253
?
? Query Match
? Best Local Similarity 2.6%; Score 68.2; DB 4; Length 543;
? Matches 188; Conservative 0; Mismatches 178; Indels 3; Gaps 1;
```

```
OY 2321 ATGAGCGTCCCTGAGACAACACAGCGGCTATGAGGGGTTCCTGGCCCTCAGCTG 2380
? | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 352 TTCATCGCCGACAGAAAATCCGAGAGATGATGCTGCAACCTGCTCAAAAGAGCC 411
? | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 2381 GAGAGCTG 2389
? | | | | |
Db 412 GGTGATCTG 420
? | | | | |
?
? RESULT 5
? PCT-US93-07923-1
? Sequence 1, Application PC/TUS9307923
? GENERAL INFORMATION:
? APPLICANT: Morimoto, Chikao
? APPLICANT: Schlossman, Stuart F.
? APPLICANT: Tanaka, Yoshiaki
? TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
? NUMBER OF SEQUENCES: 16
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson
? STREET: 225 Franklin Street
? CITY: Boston
? STATE: Massachusetts
? COUNTRY: U.S.A.
? ZIP: 02110-2804
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
? COMPUTER: IBM PS/2 Model 502 or 55SX
? OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
? SOFTWARE: WordPerfect (Version 5.0)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US93/07923
? FILING DATE: 19930819
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/934,162
? FILING DATE: 21-AUG-1992
? APPLICATION NUMBER: 07/832,211
? FILING DATE: 06-FEB-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Fraser, Janis K.
? REGISTRATION NUMBER: 34,819
? REFERENCE/DOCKET NUMBER: 00530/055002
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 542-5070
? TELEFAX: (617) 542-8906
? TELEX: 200154
?
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2924
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
?
? PCT-US93-07923-1
?
? Query Match
? Best Local Similarity 2.1%; Score 55.6; DB 5; Length 2924;
? Matches 155; Conservative 0; Mismatches 144; Indels 3; Gaps 1;
```

D _b	1910	TAGTAACTCATTATGGTCGCGGATGCAGAAAGGCCGTGTCAAGTGATGGAATACGCTG	1969
O _y	2270	GCCCCGCTAACCGTCGTGATGGGCTACGACACAGGTTACACTGAGCGCATGTGACGCT	2329
D _b	1970	GGCGCTGATATCCGGGTGGAGGTACTATGACTCAGTGTACACAGAAGCTTACATGGGTCT	2029
O _y	2330	CC 2331	
D _b	2030	CC 2031	

RESULT 6
US-08-230-491A-1

GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FELFE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 Inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5567299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2815 Base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-08-230-491A-1

Query Match	1.88;	Score 46.6;	DB 1;	Length 2015;
Best Local Similarity	54.38;	Pred. No. 0.18;		
Matches 94;	Conservative 0;	Mismatches 79;	Indels 0;	Gaps 0

[illegible]

RESULT 7
US-08-619-280A-1
; Sequence 1, Application US/08619280A

Patent No. 5767242
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
TITLE OF INVENTION: ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2815 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-619-280A-1

Query Match	1.8%;	Score 46.6;	DB 1;	Length 2815;
Best Local Similarity	54.3%;	Pred. No. 0.18;		
Matches 94;	Conservative 0;	Mismatches 79;	Indels 0;	Gaps 0

Qy	2159	GGCTTCATCGACCTGAGCCGAGTGGCCATCATAGGCTGAGGGGGCTTCCTTCG	2218
Db	2039	GGTTTCATGTGTGAAAAAAGAAATAGCCATATGGGGCTGGCTCTATGAGAGATACGTTCA	2098
Qy	2219	CTCATGGGGGCAATCCACAAGCCCAAGAGTGTCAAGGTGGCCATCGCGGGTCCCGCGTC	2278
Db	2099	TCACTGGGCCCTTGCATCTGGAACTGGTCTTTTAAATGTGTGATATAGCAAGTGGCTCCAGTTC	2158
Qy	2279	ACCGTCTGATGGGCTACGACACAGGGGTACACTGAGCGCTACATGAGCACTGCC	2331
Db	2159	TTCACGCTGGGAATTTACGCGCTGTGTCTACACAGAGAAATTCATGGGTCTCCC	2211

RESULT 8
US-08-940-391-1
Sequence 1, Application US/08940391
Patent No. 5965373
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York

```

? COUNTRY: USA
? ZIP: 10022
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: PC-DOS
? SOFTWARE: Wordperfect
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/940.391
? FILING DATE: 01-OCT-1997
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/619,280
? FILING DATE: 18-MARCH-1996
? APPLICATION NUMBER: 08/230,491
? FILING DATE: 20-APRIL-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Hanson, No. 5965373man D.
? REGISTRATION NUMBER: 30,946
? REFERENCE/DOCKET NUMBER: LUD 5330.1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 688-9200
? TELEFAX: (212) 838-3884
? INFORMATION FOR SEQ ID NO: 1.:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2815 Base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
?
? US-08-940-391-1

```

```

Query Match 1.8%; Score 46.6; DB 2; Length 2815;
Best Local Similarity 54.3%; Pred. No. 0.18;
Matches 94; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

OY 2159 GCGTTCATCGACCTGACCGGAGCTTCCATCCATGCGCTGCTCCACGCGGGGCTTCCTCG 2218
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2039 GGTTCATTGATGAAAAAAGAAATAGCCATATGGGCTGTGCTCTATGAGAGATACGTTTCA 2098

OY 2219 CTCATGGGGGCTTAATCCACAGAGCCCGAGGGTTCACAGGTGCGCATGCGGGGTGCCCCG 2278
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2099 TCATCGGCGCTTGCACTCTGGAACCTGTCTTTTCAAAATGTGTATATACAGTGGCTCCAGTC 2158

OY 2279 ACCGCTGTGAGTGCCCTACGACACAGGGTATACACTGACGCGCTTACATGAGACGTCCC 2331
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2159 TCCACTGGGAAATATTACGCGCTGTCTCTACACAGAGAAATCATGGGTCTCCC 2211

RESULT 9
US-08-858-876A-1
; Sequence 1, Application US/08858876A
; Patent No. 6022856
; GENERAL INFORMATION:
; APPLICANT: Daniel CAPUT
; APPLICANT: Pascale CHALON
; APPLICANT: Pascale FERRARA
; APPLICANT: Vito NATALIO
; TITLE OF INVENTION: Type 2 Neurotensin Receptor
; TITLE OF INVENTION: (hnt-R2)
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
; STREET: 400 Seventh Street
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

```

```

: APPLICATION NUMBER: US/08/858,876A
: FILING DATE: 19-SEP-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR 9723204
: FILING DATE: 17-MAR-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Player, William E.
: REGISTRATION NUMBER: 31,049
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1575 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 37..1266
: US-08-858-876A-1

```

```

Query Match          1.7%:  Score 45.4;  DB 3;  Length 1575;
Best Local Similarity 45.6%:  Pred. No. 0.29;
Matches 145;  Conservative 0;  Mismatches 166;  Indels 0;  Gaps 0

QY 1506 GGACAGGCTCCAGATCTGGGTCAATGAGAGAGACCAAGCTGGTACTTCCAGGGCACCA 1565
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 GCGCTGTGGGCGAAGGTGCTGTTACACCGGCGCTACGCACTCATCTGGGGCGGTGGCGGG 178
QY 1566 AGGACAGCCCGCTGGAGGACCAACCTCTACGTGTACGTATATAGGCGGGCGGAGATAG 1625
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 CGGGCAATGCGCTGTCCGTGCACGTTGTGCTAAAGCGCGGGCGGGCGGGGGCGCG 238
QY 1626 TACGCGCTACCAACAGCGCCGGCTTCTCCATATAGCTCTCCATGAGCCAGAACTTCAGATGT 1685
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 TGGCGCACACAGTGTCTCAGCGCTGCGGGGCTTCGCGGGCTGTGCTGTCTGTGTGCGGTGC 298
QY 1686 TCGTCAGCCACTACAGCAGCGGTGAGCAGCGCGCGCTGCGTGCAGCTCTTAAAGCTGAGCG 1745
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 CGGTGAGACTCTACAGCTTCGTGTGTGTTCACACTACCCCTGGGTCTTGGCGGACCTGGCGCT 358
QY 1746 GCCCGAGAGAGACCCCGTGCACAGCAAGCCCGCTTCTGGGCTAGCATGATGAGGGCAG 1805
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 GCGCGGGCTACTACTTCGTGTCAGCAGAGCTGTGCGCTTACGCCACAGGCTGTGAGCGGTGCGAG 418
QY 1806 CCAGCTGCCCC 1816
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 GCGTGAAGCGCC 429

RESULT 10
US-09-472-880-1
: Sequence 1, Application US/09472880
: Patent No. 6274333
GENERAL INFORMATION:
APPLICANT: Daniel CAPUT
           Pascale CHALON
           Pascual FERRARA
           Vito NATALIO
TITLE OF INVENTION: Type 2 Neurotensin Receptor
                  (hNT-R2)
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
STREET: 400 Seventh Street
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/472,880
FILING DATE: 28-Dec-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 9723204
FILING DATE: 17-MAR-1997

ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,049

INFORMATION FOR SEQ ID NO: 1
SEQUENCE CHARACTERISTICS:
LENGTH: 1575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 37..1266
SEQUENCE DESCRIPTION: SEQ ID NO: 1
US-09-472-880-1

```

Query Match	1.78;	Score 45.4;	DB 4;	length 1575;
Best Local Similarity	46.68;	Pred. No. 0.29;		
Matches 145; Conservative	0;	Mismatches 166;	Indels 0;	Gaps 0;

Qy	1506	GGCACGGGCTCAAGATCTGGGTCAATGAGAGACCAAGCTGGTGTACTCCAGGGCACCA	1565
Db	119	GCCCTTGGGCGCAAGTGTCTGTTTACCGGCGCTACGCACTATCTGTGGGCGCTGGGGCGGG	178
Qy	1566	AGGACACGGCCGTGAGACACCACTCTACGTGTACGTATGAGGGCGGCGGAGATCG	1623
Db	179	CGGGCAATGGCGCTGTCCGTCACAGTGTGTCTGAAGCGCGGGCCGGGCGGGGGCCCC	238
Qy	1626	TACGCTCTACACAGCCCGCGCTTCTCCATAGCTGCATGAGACCAAGACTTCGACATGT	1685
Db	239	TGGCGCACACAGTGTCTAGCTGGCGCTGGGGGCGCTGCTGTGCTGGTGGGCGGTGC	298
Qy	1686	TGCTGACGCCACTACAGCAGGTGAGCAGCGCGCCCTGCGTCACGTCTACAAGCTGAGCG	1745
Db	299	CGGTGAGGCTCTACAGTCTTGCTGTGGTTCACATACCCCTGGGTCTTTCGGCGACCGGGCT	358
Qy	1746	GCCTCCAGCAGACCCCTCTCCACACAGCCCGCGCTTCTGGGCTGACATGATGAGAGCAG	1805
Db	359	GGCGGGGCTACTACTTGTGTGCACAGACCTGTGGCGCTACGCCACAGGTGCTGAGACGTGGAG	418
Qy	1806	CCAGCTGCCCC	1816
Db	419	GCTGAGCGGCC	429

RESULT 11
 US-07-945-283-1/c
 Sequence 1, Application US/07945283
 Patent No. 5352596
 GENERAL INFORMATION:
 APPLICANT: Cheung, Andrew K.
 APPLICANT: Wesley, Ronald D.
 TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
 TITLE OF INVENTION: Involving The Ep0 and LfT Genes
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Curtis P. Ribando
 STREET: 1815 No. 5352596th University Street
 CITY: Peoria
 STATE: IL
 COUNTRY: USA

```

ZIP: 61604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4011 ext.513
TELEFAX: 309-685-4128
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8438 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
FEATURE:
NAME/KEY: CDS
LOCATION: 622..6495
FEATURE:
NAME/KEY: variation
LOCATION: replace(1099, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1267, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
NAME/KEY: variation
LOCATION: replace(7010, "g")
IS-07-945-283-1

```

Query Match	1.7%;	Score 45.4;	DB 1;	Length 8438;
Best Local Similarity	48.6%;	Pred. No. 0.44;		
Matches 124;	Conservative 0;	Mismatches 131;	Indels 0;	Gaps 0;

QY	14	GCCACCAACGGGGAGCCCAACGGGCGGACCGAGGCGAGCGAGCGCCACGAGATGACCGGCG	73
Db	5511	GCCGGGGCCCCGGGGCCCTCTGGGCTCTCTCCGCTCGCGCACTCTCTCTCCGGCGGGGCG	5452
QY	74	GCCCGCTTCCAGGTGAGAGCACTCTGTGGAGCGGGCTCTCGGAGCATCATCCAGGGCAGC	133
Db	5451	TCCCGGGCCCCGGGAGCCGCGCCGCGCCGCGGAGGAAAGCGCGCTCCACCAAC	5392
QY	134	CGCAAGTACGCGGGGCTCATTTGTCAACAGGGGGGCCCAAGACTTCATCATTTGTGAGAG	193
Db	5391	AACCACTCTCTGCTCATGGCCGAGGGGGCCCCCGGACGACGAGGGCGCTGCTACCCCG	5332
QY	194	ACGGATGAGTCTTGGGCGCCACTCCACCGCTCTACTACTTGGGAATGCCATATGGAGC	253
Db	5331	CTGGGGGAGCGCTGGCGCGGCTCGACACTCTCGGGCGAAGGGCGCGTCCGGTACGGGGCG	5272
QY	254	CGAGAGACTCCCTC	268
Db	5271	GCCGGGAGACTCCCGC	5257

RESULT 12

US-08-832-399-1

; Sequence 1, Application US/08832399

; Patent No. 6008050

; GENERAL INFORMATION:

; APPLICANT: Bergsma, Derk

; APPLICANT: Shabon, Usman

; TITLE OF INVENTION: NOVEL HUMAN NEUROTENSIN RECEPTOR TYPE 2

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Smithkline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY:

; ZIP: 19406

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/832,399

; FILING DATE: 02-Apr-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; ATTORNEY/AGENT INFORMATION:

; NAME: King, William T

; REGISTRATION NUMBER: 30,954

; REFERENCE/DOCKET NUMBER: GH50020

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-270-5515

; TELEFAX: 610-270-5090

; TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1342 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-832-399-1

Query Match

Best Local Similarity 1.7%; Score 45; DB 3; Length 1342;

Matches 144; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

Db 1508 CACGGCTCCAAGATCTGGTCAATGAGAGACCAAGCTGGTACTTCCAGGCGACCAAG 1567

Db 137 CTCTGGGCGCAAGGTGCTGTTACCGCGCTCTACGCACTATCTGGGCGCTGGGCGCGC 196

Db 1568 GACACGGCGCTGGAGCAGCACTCTACGTGTCAGTATGAGGCGCGCGAGATGTA 1627

Db 197 GGCATATGCGCTGTCGTCAGCGTGTCTGAAGCGCGGCGCGCGCGCGCGCTG 256

Db 1628 CGCCTACACAGCGCGGCTTCTCCATAGCTGCTCCATGAGCAGAACTTGACATTTT 1687

Db 257 CGCAGCAGCTGCTACAGCTGGGCGCTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTG 316

Db 1688 GTGAGCCTACAGCAGCAGTGGAGCAGCGCGCGCTGCTGACGCTCAAGCTGAGGGG 1747

Db 317 GTGAGACTCTACAGCTTCTGTGTGTCTTACCTACCGCTGGGCTTCCGCGAGCTGGGCTGC 376

Db 1748 CCCGAGCAGCAGCCCTGCAACAGCAGCCCGCTTCTGGGCTACAGATGAGAGAGCC 1807

Db 377 CGGAGCTACTACTCTGTCAGCAGAGCTGTGGCGCTACGCCAGGCTGAGAGGTGGAGCC 436

Db 1808 AGCTGCGCC 1816

Db 437 CTGAGCGCC 445

RESULT 13

; Sequence 1, Application US/09372498

; Patent No. 616182

; GENERAL INFORMATION:

; APPLICANT: Derk J. Bergsma

; APPLICANT: Usman Shabon

; TITLE OF INVENTION: NOVEL HUMAN NEUROTENSIN RECEPTOR TYPE 2

; FILE REFERENCE: GH-50020-1

; CURRENT APPLICATION NUMBER: US/09/372,498

; CURRENT FILING DATE: 1999-08-11

; PRIOR APPLICATION NUMBER: 08/832,399

; PRIOR FILING DATE: 1997-04-02

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 1342

; TYPE: DNA

; ORGANISM: HOMO SAPIENS

US-09-372-498-1

Query Match

Best Local Similarity 1.7%; Score 45; DB 4; Length 1342;

Matches 144; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

Db 1508 CACGGCTCCAAGATCTGGTCAATGAGAGACCAAGCTGGTACTTCCAGGCGACCAAG 1567

Db 137 CTCTGGGCGCAAGGTGCTGTTACCGCGCTCTACGCACTATCTGGGCGCTGGGCGCGC 196

Db 1568 GACACGGCGCTGGAGCAGCACTCTACGTGTCAGTATGAGGCGCGCGAGATGTA 1627

Db 197 GGCATATGCGCTGTCGTCAGCGTGTCTGAAGCGCGGCGCGCGCGCGCGCTG 256

Db 1628 CGCCTACACAGCGCGGCTTCTCCATAGCTGCTCCATGAGCAGAACTTGACATTTT 1687

Db 257 CGCAGCAGCTGCTACAGCTGGGCGCTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTG 316

Db 1688 GTGAGCCTACAGCAGCAGTGGAGCAGCGCGCGCTGCTGACGCTTCAAGCTGAGGGC 1747

Db 317 GTGAGACTCTACAGCTTCTGTGTGTCTTACCTACCGCTGGGCTTCCGCGAGCTGGGCTGC 376

Db 1748 CCCGAGCAGCAGCCCTGCAACAGCAGCCCGCTTCTGGGCTACAGATGAGAGAGCC 1807

Db 377 CGGAGCTACTACTCTGTCAGCAGAGCTGTGGCGCTACGCCAGGCTGAGAGGTGGAGCC 436

Db 1808 AGCTGCGCC 1816

Db 437 CTGAGCGCC 445

RESULT 14

US-09-221-017B-646/c

; Sequence 646, Application US/09221017B

; Patent No. 644799

; GENERAL INFORMATION:

; APPLICANT: ROSS, Bruce C.

; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/221,017B

```

1      FILING DATE: 23-DEC-1998
2      CLASSIFICATION:
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER: PP1182
5      FILING DATE: 31-DEC-1997
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER: PP1546
8      FILING DATE: 30-JAN-1998
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER: PP2911
11     FILING DATE: 09-APR-1998
12     PRIOR APPLICATION DATA:
13     APPLICATION NUMBER: PCT/AU98/01023
14     FILING DATE: 10-DEC-1998
15     ATTORNEY/AGENT INFORMATION:
16     NAME: Monroy, Gladys H
17     REGISTRATION NUMBER: 32,430
18     REFERENCE/DOCKET NUMBER: 27340-20021.00
19     TELECOMMUNICATION INFORMATION:
20     TELEPHONE: 650-813-5600
21     TELEFAX: 650-494-0792
22     TELEX: 706141
23     INFORMATION FOR SEQ ID NO: 646:
24     SEQUENCE CHARACTERISTICS:
25     LENGTH: 657 base pairs
26     TYPE: nucleic acid
27     STRANDEDNESS: double
28     TOPOLOGY: circular
29     MOLECULE TYPE: DNA (genomic)
30     HYPOTHEICAL: NO
31     ANTI-SENSE: UNKNOWN
32     ORIGINAL SOURCE:
33     ORGANISM: PORYPHYROMONAS GINGIVALIS
34     FEATURE:
35     NAME/KEY: misc_feature
36     LOCATION: 1...657
37     OS-09-221-017B-646

```

```

1  APPLICANT:  Launis, Karen L.
2  APPLICANT:  Rothstein, Steven J.
3  APPLICANT:  Bowman, Cindy G.
4  APPLICANT:  Dawson, John L.
5  APPLICANT:  Dunder, Erik M.
6  APPLICANT:  Pace, Gary M.
7  APPLICANT:  Suttie, Janet L.
8  TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
9  TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
10 NUMBER OF SEQUENCES:  94
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE:  CIBA-GEIGY Corporation
13 STREET:  7 Skyline Drive
14 CITY:  Hawthorne
15 STATE:  New York
16 COUNTRY:  USA
17 ZIP:  10532
18
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE:  Floppy disk
21 COMPUTER:  IBM PC compatible
22 OPERATING SYSTEM:  PC-DOS/MS-DOS
23 SOFTWARE:  PatentIn Release #1.0, Version #1.30B
24
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER:  US/07/951,715A
27 FILING DATE:  25-SEP-1992
28 CLASSIFICATION:  800
29
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER:  US 07/772,027
32 FILING DATE:  04-OCT-1991
33
34 ATTORNEY/AGENT INFORMATION:
35 NAME:  Spull, W. Murray
36 REGISTRATION NUMBER:  32,943
37 REFERENCE/DOCKET NUMBER:  S-18805/A/CGC 1577/CIP
38
39 TELECOMMUNICATION INFORMATION:
40 TELEPHONE:  (919)541-8615
41 TELEFAX:  (919)541-8689
42
43 INFORMATION FOR SEQ ID NO:  6:
44 SEQUENCE CHARACTERISTICS:
45     LENGTH:  3624 base pairs
46     TYPE:  nucleic acid
47     STRANDEDNESS:  single
48     TOPOLOGY:  linear
49     MOLECULE TYPE:  other nucleic acid
50     DESCRIPTION:  /desc = "Synthetic DNA"
51     HYPOTHETICAL:  NO
52     FEATURE:
53         NAME/KEY:  CDS
54         LOCATION:  1..3621
55         OTHER INFORMATION:  /product= "Full-length, maize
56                             optimized cDNA"
57         OTHER INFORMATION:  /note= "Disclosed in Figure 6."
58
59 US-07-951-715A-6
60
61 Query Match          1.7%;  Score 44.4;  DB 1;  Length 3624;
62 Best Local Similarity 48.1%;  Pred. NO. 0.6;
63 Matches 126;  Conservative 0;  Mismatches 136;  Indels 0;  Gaps 0;
64
65 1509  ACGGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGCGACCAAG 1568
66      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 Db 1403  ACCGCGACCAACACCTATCGGCCCAACCGCATCACCACCATGCCATGTGGAAGGCCAGG 1462
68
69 1569  ACACGCCCTGTGAGACCAACCTCTACGTGTGAGCTATGAGGCGGCGCGAGAGTGTAC 1628
70      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 Db 1463  AGCTGCCCCAGGAGCAACACCGTGTGTGCGCGCGCGCGCTTACACCGGCGGACATCTGCG 1522
72
73 1629  GCCTCAGCAAGCGCGGCTTCCATAGCTGCTGCATGAGCCAGAACATTGACATGTTGCG 1688
74      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 Db 1523  GCCGCAACCAACACCGCGGCGCTTCCGCGCCCATCGGCTGACCGTGAAGCGGCCCTGACCC 1582
76
77 1689  TCAGCCTACTACAGACGCTGAGACAGCGCGCGCTGCGTGTGACACTTCAAGAGCTGAGCGGCC 1748
78      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79 Db 1583  AGCGCTACCGGATCGGCTTCCGCTTACGCGCAGACACCGTGGAGCTTGGACTTCTTCTGTGAGCC 1642

```


QY 1749 CCGACGACGACCCCTGCACAA 1770
| | | | | | | | | |
Db 1643 GCGGCGGACCCCGTGACAA 1664

Search completed: December 12, 2002, 10:38:49
Job time : 121 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame-plus_n2p model

Run on: December 12, 2002, 10:38:52 : Search time 21.5 Seconds
(without alignments)
10097.074 Million cell updates/sec

Title: US-09-976-674-4
Perfect score: 4795
Sequence: 1 caagcttaccatgacccacca.....tctgagcgccgagatccg 2617

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame-n2p.model -DEV=xlh
-Q=/c9n2_1/USPRO.spool/US09976674/runatc_04122002_162359_5975/app-query.fasta.1.2759
-DB=Swissprot.40 -QFMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=ppct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ppct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09976674 -CCGN_1_1_1_1 -etunal_04122002_162359_5975 -NCP=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Swissprot.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	505.5	10.5	765	1	DP4_BOVIN
2	498.5	10.4	767	1	DP4_BOVIN
3	489.5	10.2	766	1	DP4_HUMAN
4	486	10.1	765	1	DP4_FELCA
5	481	10.0	760	1	DP4_MOUSE
6	452	9.4	865	1	DP6_HUMAN
7	447.5	9.3	859	1	DP6_RAT
8	442.5	9.2	863	1	DP6_BOVIN
9	438.5	9.1	761	1	SEPR_MOUSE
10	432.5	9.0	818	1	DAP2_YEAST
11	426	8.9	760	1	SEPR_HUMAN
12	417	8.7	931	1	DAP1_YEAST
13	359.5	7.5	1367	1	AMYH_YEAST
14	279.5	5.8	5179	1	MCC2_HUMAN
15	265.5	5.5	660	1	YHL1_EBV
16	263	5.5	660	1	YHL1_EBV
17	251.5	5.2	1763	1	CA24_ASCSU
18	250.5	5.2	1324	1	IRS2_HUMAN

19	249.5	5.2	1690	1	CA44_HUMAN	P53420 homo sapien
20	249	5.2	657	1	YUXL_BACSU	P39839 bacillus su
21	247.5	5.2	1603	1	CA1F_HUMAN	O07092 homo sapien
22	244.5	5.1	1251	1	Y003_CAEEL	O09350 caenorhabdi
23	242.5	5.1	1262	1	CA13_CHICK	P12105 gallus gall
24	241.5	5.0	1199	1	N121_RAT	P52591 rattus norv
25	239.5	5.0	1783	1	RAA3_CHLRE	O9fec4 chlamydomon
26	238.5	4.9	1464	1	CA13_MOUSE	P08121 mus musculu
27	238.5	4.9	1466	1	CA13_HUMAN	P02461 homo sapien
28	238	4.9	1262	1	CA13_CHICK	P12105 gallus gall
29	238	4.9	1459	1	CA12_MOUSE	P28481 mus musculu
30	237.5	4.9	1049	1	CA13_BOVIN	P04258 bos taurus
31	237.5	5.0	1183	1	DRPL_RAT	P54258 rattus norv
32	235.5	4.9	1366	1	CA21_HUMAN	P08123 homo sapien
33	235	4.9	1461	1	IE18_PRIVF	P11675 pseudorabie
34	235	4.9	1464	1	CA11_HUMAN	P02452 homo sapien
35	232.5	4.8	1460	1	CA11_CANFA	O9xsj7 canis famli
36	232	4.8	1185	1	DRPL_HUMAN	P54259 homo sapien
37	230	4.8	2944	1	CA17_HUMAN	O02388 homo sapien
38	229.5	4.8	1372	1	CA21_RAT	P02456 rattus norv
39	229.5	4.8	1453	1	CA11_CHICK	P02457 gallus gall
40	229	4.7	1418	1	CA12_HUMAN	P02458 homo sapien
41	229	4.7	1453	1	CA11_MOUSE	P11087 mus musculu
42	228	4.8	671	1	CA11_RAT	P02454 rattus norv
43	228	4.8	1446	1	IE18_PRIVKA	P33479 pseudorabie
44	227.5	4.7	1255	1	MUCL_HUMAN	P15941 h muctin 1 p
45	227.5	4.7	1453	1	CA11_CHICK	P02457 gallus gall

ALIGNMENTS

RESULT 1
DP4_BOVIN
ID P81425; Q8WMC8;
AC DP4_BOVIN STANDARD; PRT: 765 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (adenosine deaminase complexing protein) (ADCP-I)
DN (Activation molecule 3) (ACT3) (MC10).
GB DP4 OR CD26.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
OX [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Lymphocytes;
RX PubMed-12073152;
RA Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J., Bohach G.A.;
RT "Molecular characterization of bovine CD26 upregulated by a staphylococcal superantigen".
RT Immunogenetics 54:216-220(2002).
RL [2]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC TISSUE=Thymus;
RX PubMed-11981836;
RA Gliddon D.R., Howard C.J.;
RT "CD26 is expressed on a restricted subpopulation of dendritic cells in vivo".
RT Eur. J. Immunol. 32:1472-1481(2002).
RL [3]
RN SEQUENCE OF 1-24.
RC TISSUE=T-cell;
RX PubMed-11598101;
RA Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K., Naessens J., Bohach G.A.;
RT "Identity of activation molecule 3 on superantigen-stimulated bovine cells is CD26".
RT Infect. Immun. 69:7190-7193(2001).
RL

FN	[4]	SEQUENCE OF 537-546.	
RP	TISSUE=Kidney:		
RC	MEDLINE=98293306; PubMed=9629661;		
RX	Ben-Shooshan I., Parola A.H.;		
RA	"The Cp-I subunit of adenosine deaminase complexing protein from calf		
RT	kidney is identical to human, mouse, and rat dipeptidyl peptidase		
RT	IV.";		
RL	Comp. Biochem. Physiol. 119B:289-292(1998).		
CC	-I- FUNCTION: Removes N-terminal dipeptides sequentially from		
CC	polypeptides having unsubstituted N-termini provided that the		
CC	penultimate residue is proline. Binds and regulates the activity		
CC	of ADA.		
CC	-I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb- -		
CC	Xcc, from a polypeptide, preferentially when Xbb is Pro, provided		
CC	Xcc is neither Pro nor hydroxyproline.		
CC	-I- SUBUNIT: Homodimer.		
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in		
CC	a soluble form (By similarity).		
CC	-I- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and		
CC	several immune system tissues.		
CC	-I- PIM: The soluble form (SDPP) derives from the membrane form (MDPP)		
CC	by proteolytic processing (By similarity).		
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage of and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@sib-sib.ch).		
CC	-----		
DR	EMBL; AF461806; AAL67836.1; -		
DR	EMBL; AY056834; AAL23628.1; -		
DR	MEROPS; S09.003; -		
DR	InterPro: IPR002469; DPPV.N term.		
DR	InterPro: IPR001375; Peptidase_S9.		
DR	InterPro: IPR002471; Prol_endopep_ser.		
DR	InterPro: IPR000379; Ser_estrs_site.		
DR	Pfam; PF00930; DPPV.N term; 1.		
DR	Pfam; PF00326; Peptidase_S9; 1.		
DR	PROSITE; PS00708; PRO_ENDOPEP_SER; 1.		
KM	Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;		
DR	Transmembrane; Glycoprotein; Signal-anchor		
FT	CHAIN	1	765
FT			
FT	CHAIN	38	765
FT			
FT	DOMAIN	1	69
FT			
FT	TRANSMEM	7	26
FT			
FT	DOMAIN	30	765
FT	ACT_SITE	629	629
FT	ACT_SITE	707	707
FT	ACT_SITE	739	739
FT	CARBOHYD	84	91
FT	CARBOHYD	91	91
FT	CARBOHYD	149	149
FT	CARBOHYD	218	218
FT	CARBOHYD	228	228
FT	CARBOHYD	271	271
FT	CARBOHYD	280	280
FT	CARBOHYD	320	320
FT	CARBOHYD	392	392
FT	CARBOHYD	495	495
FT	CARBOHYD	684	684
FT	SEQUENCE	765 AA; 88369 MW; E22155421F43E116 CRC64;	

Alignment Scores:	2.41e-
Pred. No.:	505.50
Score:	39.15%
Percent Similarity:	

```

Length: 765
Matches: 180
Conservative: 98

```

Best Local Similarity:	25.35%	Mismatches:	269
Query Match:	10.54%	Indels:	163
DB:	1	Gaps:	28
US-09-976-674-4 (1-2617) x DPP4_BOVIN (1-765)			
QY 614 TCCTTCATCATATACAGCCAGCTGTGGTGGCCCAATCATCGAGACAGCGAGCGCGG	673		
Db 164 AletayValtrpmsnasbpietleyValylsasngluProasnsrProserlalnrg	183		
QY 674 CTGACCTTTCACCAACAGGCTTTATCCAAATGTCCTGGATGCCACCCAGCTGCGGGTGTG	733		
Db 184 IlethrTrp-----ThrglylsylsaspaValIleTyrAsn-----Glyle	197		
QY 734 GCCACCTTCGTCATACAGGAAGAG---TTCGACCGCTTCACAGGGTACTGCTGGTGGCCCC	790		
Db 198 ThrasprValylrgngluGluValPheSerAlatYSerAlaLeutrPrtrserPro	217		
QY 791 ACACGCTTCGCGGAAGGTTCAGAGGGCCTCAGACGCTCGATCCGTATGAGCAATGC	850		
Db 218 Asnserrhrhe-----LeuAlatYrAlaGlnhe	227		
QY 851 GATGAGTCCAGAGTGGAGTCACTAC-----GTCCCTTCCTGCGCTACAGAA	901		
Db 228 AsnaspHrGluValProleuIleGluTyrSerPheTyrSerAspIserLeuIndTyr	247		
QY 902 AGGAAGAGACATCGTATCGTACCCACAGACAGCAGCAATCCCAATGTTCTCTG	961		
Db 248 ProlysrhrValylslierProTyrProlysrAlaGlyAlaValasnrProthrlle-----	265		
QY 962 AAATGCGCTAGATTCACAGACTGCACAGCCAGGCGAGATGCTTCGACCCAGAGAGAG	1021		
Db 266 LyshePheValValasnlIeserSerLeuSerProasnlIeasAlaThrSerlIn	285		
QY 1022 CTGTGTGACGCTTCAGCTGCTGTGTCCGAAGGTGAGTACATACCGACGGCGGGTGG	1081		
Db 286 IleValProProGlysValIleuIleGlyAsp---HisTyrLeuCyAspValThrTrp	304		
QY 1082 ACCGGATGSCAAATATACGCTTGCGCCATGTTCTTGAGACGGCCCAAGCATGGCTCCAG	1141		
Db 305 ValThrGluIuArgIleSer-----LeuGlnTrpLeu---	315		
QY 1142 CTGTCTCTCTCCCGCGCGCCGTTCATTCGAGACAGAGAATAGAGAGAGCGGCTA	1201		
Db 315 -----	315		
QY 1202 GCGTCTCCAGAGCTGTCCCGAGAAATGTCCAGCCGTATGTGTG---	1246		
Db 316 -----ArgArgIleGlnasnrTyrSerIleMetAspIlecyasp	328		
QY 1247 TAGCAGAGAGTCAACCAACGCTGTGCATATGTTCATGACATCTTTATCCCTTCCCCAA	1306		
Db 329 TyrAspArgSerThrGlyArgTrpIleSer-----	338		
QY 1307 TCAGAGGGAAGAGACGAGCTCTCTTTCGCGGCCAATGAATGCAAGACGGCTTGTG	1366		
Db 339 -----SerValGlyArgGln	343		
QY 1367 CATTTGTACAAGTCAACCGCGCTTTTAAATCCAGAGGCTACAGATGGAGTGAGCCCTTC	1426		
Db 344 HisIleIuIleserThrThr-----GlyTrpValGlyArgPhe	356		
QY 1427 AGCCCCGGGGAAGATGAATTT-----AAGTGGCCCATTTAG	1482		
Db 357 ArgProAlaGluProHisPheThrSerAspGlyAsnSerPheTyrLysIleIleSerAsn	376		
QY 1463 GAAGAG-----ATTGCTCTG	1477		
Db 377 GluGluGluTyrLysHisIleCyHisPheGlnThrAspLysArgAsnCysThrPheIle	396		
QY 1478 ACAAGCGGTAAATGGAGAGTTTGGCGAGGACAGGCTCCAGATGTGGGTCAATGAGAG	1537		
Db 397 ThrLysIleValatPrpIuValIleGly-----IleGluAlaLeuThrSer	411		

Db 96 AspIuIhegIyHisSerIleasnAspTySerIleSerProaspGlyIuPheIleu 115
 QY 470 CAGCCGCAAGAACAGCCCTTCCACTGCGCAGCGCGCAAGAACGGTTATGTC 529
 Db 116 LeuGIuTyrsnTyValLySgIuTrpHisSerTyThrAlaSerTyAspIleTy 135
 QY 530 CCTTGAACCGCGTGAATCAAGCCAGTGTACAGGCGCCGAGTGGACCAAAATC 589
 Db 136 AspleuasnTyAspGlnIleuIleThrGluuArgIleProasnAsnThrGlnTrpVal 155
 QY 590 TGCCCTCCGACCCCTGCCCTTC--TTCTCTTCATCAATAACAGCAGCTGTGGTGCC 646
 Db 156 ThrTrpSerProValGlyHisTyLeuAlaTyValTrpAsnAsnAspIleTyValLyS 175
 QY 647 AACATTCAGACAGCGGAGAGCGGCGGTGACCTTTCGCCCAAGTTTATCCAAATGTC 706
 Db 176 IleGIuProasnLeuProSerTyTrpArgIleThrTrp-----ThrGlyLySgIuAspIle 193
 QY 707 CTGATGACCCCAAGTGTGCGGTGTGGCCACCTTCGATACAGGAAG--TTGCAC 763
 Db 194 IleTyrsn-----GlyIleThrAspTrpValTyrgIuGluValPheSer 209
 QY 764 CGCTTCACTGCTACTGTGTGTGCCACAGCTCTGGGAAGGTTCAGAGGCTTCAG 823
 Db 210 AlatySerAlaLeuTrpTrpSerProasnGlyThrPhe----- 222
 QY 824 ACCTGCGAATCTCTGTAGAGAAAGTGCATGATCCAGATCGAGATTCATTCAC----- 877
 Db 223 -----LeuAlaTyraGlnPheasnAspThrGluValProleuIleGlyTySer 239
 QY 878 ---GTCCCTCTCCTGCGCTGAGAAAAGAGAGAGAGCTGTATCGGTACCCAGGACA 934
 Db 240 PheTySerAspGlnSerLeuGlnTyTrpProLyThrValAlaGValProTyProLyAla 259
 QY 935 GGCAGCAAGATCCCAAGATTCCTTGAACCTGCTGATTCAGACTGACAGCCAGGCG 994
 Db 260 GlyAlaValAsnProThrVal-----LysPheValValAsnThrAspSerLeuSer 277
 QY 995 AAGATCTCTTCGACCCAGAGAGAGAGAGTGTGACGCCCTTCAGCTGTCCCGAAG 1054
 Db 278 SerValThrAsnAlaThrSerIleGlnIleThrAlaProAlaSerMetLeuIleGlyAsp 297
 QY 1055 GTGAGTACATCCGCGCAGGCGGTGAGACCGGATGGCAATAGCGCTGGGCCATGTC 1114
 Db 298 ---HisTyLeuCyAspValThrTrpAlaThrGlnGluArgIleSer----- 312
 QY 1115 CTGACCGGCGCCAGAGTGGCTCCAGCTGCTCTCTCCCGCGGCGCTTCATCCCG 1174
 Db 313 -----LeuGlnTrpLeu----- 316
 QY 1175 AGCACAAGATGAGAGAGAGCGGTAGCCTCTCCAGAGCTGTCCCGAATGTCCAG 1234
 Db 317 -----ArgArgIleGln 320
 QY 1235 CCGATGTGCTG-----TAGAGAGAGTACCAACGCTGTGATCAATGTT 1279
 Db 321 AsnTySerValMetAspIleCyAspTyTrpAspLeuSerGlyArgTrp----- 337
 QY 1280 CATGACATCTTCTATCCCTTCCCAATCAAGAGGAGAGAGAGCTGTCTTCGCGC 1339
 Db 337 ----- 337
 QY 1340 GCCAATAATGCAAGACCGGCTTGTCCATTTGTACAAGTCCACCGCGTTTAAATCC 1399
 Db 338 -----AsnCySLeuValAlaArgGlnHisIleGluMetSerThrThr----- 351
 QY 1400 CAGGGTACGATGGAGTGAAGCCTTCAGCCCGGGGAAGATGA----- 1444
 Db 352 -----GlyTrpValGlyArgPheArgProSerGluProHisPheThrLeuAspGly 368
 QY 1445 -----TTTAAAG 1450
 Db 1445 -----TTTAAAG 1450

Db 369 AsnSerPheTyLySleIleSerAsnGluGlyTyArgGlnHisIleCyTyTrpGln 388
 QY 1451 TGCCCATTAAGAAAGATGCTGTGCACACCGGTGAATGGAGCTTTGGCAGGAC 1510
 Db 389 IleAspLySAspCySThrPheIleThrLySgIuThrTrpGluValIleGlyIleGlu 408
 QY 1511 GGCITCAAGATCTGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGCAACCAAGAGAC 1570
 Db 409 AlaLeu-----ThrsrAsp 413
 QY 1571 AGCGCGTGAAGACACCACTCTACGTGTGACATAAGAGGCGCGCG----- 1618
 Db 414 -----TyLeuTyTrIleSerAsnGluTyLySgIuMetProGly 428
 QY 1619 ---GAGATCGTACGCTTACCAACCGCGCTTTCCTTCCATGCTGCTCATAGCCAGAAC 1675
 Db 429 ArgAsnLeuTyLySleIleGlnLeuSerAspTyThrLySValThrCySLeuSerCySglu 448
 QY 1676 TTGCACATGTCGTCAGCCTACAGCAGCGTGAAGAGAGCGCGCGCTGCTGACGCTAC 1735
 Db 449 LeuAsnProGluuArgCySglnTyTrpTySerValSerPheSerLySgIuAlaLyTyTr 468
 QY 1736 AAGCTG-----AGCGCGCC----- 1750
 Db 469 GlnLeuArgCySserGlyProGlyLeuProLeuTyThrLeuHisSerSerValAsnAsp 488
 QY 1751 -----GAGCAGACCCCTTCGCAAGACAGCCCGCTTGGCGT 1789
 Db 489 LySgIuLeuArgValIleuGluAspAsnSerAlaLeuAspLyS----- 502
 QY 1790 AGCATGATGAGGAGCGCAGCGTGCCTCCGATATGTCCTCCAGATGTCATTCATTC 1849
 Db 503 ---MetLeuGlnAsnValGln-----MetProSerLySLeuAspPhe 516
 QY 1850 CACACGCGCTCGATGTGCGGCTC--TAGCGATGATCTACAAAGCCAGCCTTGCAG 1906
 Db 517 IleIleLeuAsnGluThrLySThrPheTrpTyrgIuMetIleLeuProHis--PheAsp 535
 QY 1907 CCAGGAAAGAAAGACCCCAACCGCTCTTGTATGTGAGGCGCCAGAGTGCAGCTGTG 1966
 Db 536 LySerLySlyTyTrpProLeuLeuAspValTyraGlyProCySserGlnLySAla 555
 QY 1967 AATACCTCTTCAAAAGCATCAAGTACTGGCGCTCAAC-----ACATGGCCCTC 2017
 Db 556 AspThrValPhe-----ArgLeuAsnTrpAlaThrTyLeuAlaSer 569
 QY 2018 CTGGGCTACGCGGTGTGTG--ATTGACGCGAGGCGCTCTCTCAGCAGAGGCTTCG 2074
 Db 570 ThrGluAsnIleIleValAlaSerPheAspGlyArgGlySerGlyTyrgIuGlyAspLyS 589
 QY 2075 TTCGAAAGGCGCCCTGAAAAACCAATGGGCCAGGTGAGATCGAGACAGAGTGAAGGCG 2134
 Db 590 IleMetHisAlaIleAsnArgArgLeuGlyThrPheGluValGluAspGlnIleGluAla 609
 QY 2135 CTG---CAGTTCGTGGCGGAGAGATAGCTTCAATCGACGAGCGGTGCGCATCAT 2191
 Db 610 AlaArgGlnPhe-----SerLySMetGlyPheValAspAsnLySArgIleAlaIleTrp 627
 QY 2192 GCGTGTCTTACGAGGCGCTCTCTCTGCTCATGGGCTAATCCAAAGCCCGAGTGTTC 2251
 Db 628 GlyTrpSerTyrgIuGlyTyValThrSerMetValIleuGlySerGlySerGlyAlaPhe 647
 QY 2252 AAGTGGCCATCGCGGCGTGCCTCCGCTACCGCTGTGATGCGCTACGACACAGGTAAPT 2311
 Db 648 LyScySgIyIleAlaValAlaProValSerArgTrpGlyTyTyAspSerValTyThr 667
 QY 2312 GAGCGCTACGAGCAGCT-----CCTGAGAACCAAGCAAGCGGTATGAGAGCGGCTCC 2365
 Db 668 GlnArgTyMetGlyLeuProThrProGluAspAsnLeuAspHisTyArgAsnSerThr 687
 QY 2366 GTGGCGCTGACAGTGAAGAGCTGCCAATGAGCCCAACCGCTTATCTTACACAGCGC 2425
 Db 688 ValMetSerArgAlaGluAsnPhe-----LysGlnValGlyTyLeuLeuIleHisGly 705

```

QY 2426 TTCCTGGACGAAGCTGCTTTTCCACCAAACTCCTCGTCCCACTGATCCGA 2485
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 706 Thrlaasprasnvalhlnsphegnlgnlserlalnleeserlysalalevalasp 725
QY 2486 GCAGGGAACCTTACCACTCCAGATCCACCCAGAGAGACAGACTATTCGTCGCC 2545
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 726 Valglyalasphegnlnalameitrrptyrthrapslnsphtsglylealaser 745
QY 2546 GATCGGGCGAGCACTAGAGTACCTGCTGCTTCTTACAGGA 2593
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 746 Thrlahisgnlnhstlerythrlnsmetserhspheileysgn 761

RESULT 4
DP4 FELCA STANDARD; PRT; 765 AA.
AC 09N217;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
GN antigen CD26).
OS DPP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RA MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RA "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue.";
RL Immunogenetics 50:366-368(1999).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -1- PFM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB023952; BAA92344.1; -.
CC MEROPS: S09.003; -.
DR InterPro: IPR002469; DPPV_N-term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; ProL_endopep_ser.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00930; DPPV_N-term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT CHAIN 38 765 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).

```

```

FT TRANSMEM 7 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 30 765 (POTENTIAL).
FT ACT_SITE 629 629 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 765 AA; 88213 MW; 3EFC98A22B175D9 CRC64;

```

```

Alignment Scores:
Pred. No.: 2,28e-17 Length: 765
Score: 486.00 Matches: 200
Percent Similarity: 38.518 Conservative: 135
Best Local Similarity: 22.998 Mismatches: 331
Query Match: 10.148 Indels: 204
DB: Gaps: 33

```

US-09-976-674-4 (1-2617) x DPP4_FELCA (1-765)

```

QY 185 GTCCAGAGAGAGATGATGTGGGCCCCACTCCACCGCTTACTGAGTAATCCGA 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28 Leuasnlysglnspaspaalalalalaspserargrghrthrleu----- 44
QY 245 TATGGCAGCGGAGAGAACTCCCTCTACTGAGATTCCCAAGAAAGTCGGAAAGAG 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 45 -----Thrasprtleuysantrhrhealval 54
QY 305 GCTCTGCTGCTCCTGCTGCTGGAAGCAGANTCTGATCATTTCCAGGACGCCACAT 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 Lysphetrserleuairtrp-----Valseraphisasp 66
QY 365 GGGGTACTCTCTGGGAGGAGAGCTGCTGAGGACCGGAACGCTGGGGCTTCGGC 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 TytleutrylsglnspaspaalalaleuLeuPhesnaIaeglutyrgly----- 83
QY 425 ATGACCTCCACACATTCACACAGGAGGCGCTCTCTCTCCAGCGACACAGC 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 -----AsnserSerleheleu-----Gluasner 92
QY 485 CTCTTCACCTGTCGCGACGGCGGCAAGAACGGCTTCATGCTCCCT----- 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93 ThrPhesplurheglunhsSerlleasnaSprlySerValserProasplglnPhe 112
QY 532 ----- 532
Db 113 lleuleuIuIyrasnTyValysglnTrprghisSerTyThrIalaserTyasp 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 533 -----ATGAACCGCTGGAATCAAGACAGCGAGCGCCCGGATGAGACCC 583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 IleTyAspIeunlsnysalaglnleuIleThrglnlulyslleProasnaInrln 132
QY 584 AAATGTGCGCCCTCGACCGCTGCTTC-----TTCCTTCATCAATGAAGACGCTGG 640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 153 TrpIleThrTrpSerProglunhslyslsleuAlatyValItrprlyasnaValTy 172
QY 641 GTGGCCAACTGACAGACAGGCGGAGCGGCTGACTTCGCAACCAAGTTATCC 700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 VallyAsnIuIProasnsrSerSerhlsArgIleThrtrp-----ThrgIuglu 190
QY 701 AATGTCTGATGATACCCCAAGCTGGCGGCTGGCCACTTCGTCATAGCAAGAG-- 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 AsnaIleTyasn-----GlyllealaspIryValTyrgIugluIle 206

```


QY 536 AAACGCTGGAATCAAGACCCAGTCTGACAGCGCCGGATGAGACCCCAAAATCTGCCCT 595
 Db 132 AsnLysArgGlnLeuIleThiGlnLysIleProAsnAsnThrGlnTrpIleThiTrp 151
 QY 596 GCCGACCCCTGCCCTTC---TTCCTCTCATCAATAAAGCAGCATGTGGGTGGCCAAACATC 652
 Db 152 SerProGlnGlnHisLysLeuAlaIleValTrpValTrpLysAsnAspIleTyrVal---LysVal 170
 QY 653 GAG-----ACAGCGCAGGACGGCGGCTACCTTC 682
 Db 171 GlnProHisLeuProSerHisArgIleThrSerThrGlnGlnIle----- 185
 QY 683 TGGCACAAGGTTATCCAAATGCTCGGATGACCCCAAGTGTGGGGGTGGCCACCTTC 742
 Db 186 -----AsnValIleTyrAsn-----GlyIleThrAspTrp 195
 QY 743 GTCATACAGAGAGAG---TTCGACCGGCTTCACTGGTACTGTGGTGGCCACAGCCTTC 799
 Db 196 ValTyrGlnGlnGlnValPheGlyAlaIleTyrSerAlaIleTyrTrpSerProAsnAsnThr 215
 QY 800 TGGAGAGTTTCAGAGGCGCTTCAGACGCTGCGAAATCCTGTATGAGAGATGATGATGCTC 859
 Db 216 Phe-----LeuAlaTyrAlaGlnPheAsnAspThr 225
 QY 860 GAGGTGAGGTCATTTCAC-----GTCCCTCTCTCGGGCTAGAGAAAGAGAGACG 910
 Db 226 GlyAlaProLeuIleGlnTyrSerPheTyrSerAspGlnSerLeuGlnTyrProLysThr 245
 QY 911 GACTCGATATCGGATCCCGCAGCAGGCGCAGCAAGATTCAGATTGCTTGAACCTGGCT 970
 Db 246 ValTrpIleProTyrTrpLysAlaGlyAlaValAsnProThrValLysPhePheIleVal 265
 QY 971 GAGTTCAGACTACAGCCAGGCGCAAGATGCTGTGACCCAGAGAGAGAGAGCTGGTCAG 1030
 Db 266 AsnIleAspSerLeuSerSer----- 272
 QY 1031 CCCCTTACGCTGCGTGTCCCGAAGGTGAGTACATCCGCGAGGCGGCTGACCCGGGAT 1090
 Db 273 -----SerSerSerAlaAlaIleProIleGlnIleProAlaProAlaSerValAlaIleGly 290
 QY 1091 GGCAGAAATAC-----GCGTGGCGCATGTTCTCTGACGCGGCCACAGCTAGGCTC 1138
 Db 291 AspHisTyrIleLeuCysAspValValTrpAlaThrGlnGlnAlaGlyIleSerLeuGlnTrpLeu 310
 QY 1139 CAGCTCGTCTCTCCCCCGGCGCTGTTCATCCCGACACAGAAATGAGAGCAGCGG 1198
 Db 310 ----- 310
 QY 1199 CTAGCCTTGGCCAGAGTGTCCCGAAGTCCAGCGGCTATGTGCTG----- 1246
 Db 311 -----ArgArgIleGlnAsnTyrSerValMetAlaIleCys 322
 QY 1247 ---TACGAGAGAGTC----- 1258
 Db 323 AspTyrAspLysIleAsnLeuThrTrpAsnLysProSerGlnGlnGlnHisValGlnMet 342
 QY 1259 ---ACCAAGCTGTGCATGATGTCATGACATCTTATCCCTTCCSCCAATCAGAGGGA 1315
 Db 343 SerThrThrGlnTyrValGlyArgPheArgProAlaGlnProHisPheThrSerAspGly 362
 QY 1316 GAGGACGAGCTGTGCTTCTCCGGGCGCAATGATGACAGACCGGCTC-----TGC 1366
 Db 363 SerSer-----PheTyrLysIleIleSerAspLysAspGlyTyrLysHisIleCys 379
 QY 1367 CATTTGTACAAGTACACCGCGCTTTTAAATCCAGGCGTACGATTGAGTGAAGCCTTTC 1426
 Db 380 HisPheProLys----- 383
 QY 1427 AGCCCCGGGGAAGATTAATTAAGTCCCATTAAGAGAGATTCCTGTACACGCGGT 1486
 Db 384 -----AspLysLysAspCysThrPhe-----IleThrLysGly 394

QY 1487 GAATGGAGAGTTTTGGCGAGGCGACGCGCTCCAGATCTGGTCAATGAGAGACCAAGCTG 1546
 Db 395 AlaTrpGlnValIleSer-----IleGlnAlaLeuThrSerAspTyrLeu 409
 QY 1547 GTGTACTTCCAGGGCACAC---AAGACACGCGCTGGAGACACACCTTACGTGGTCAAGC 1603
 Db 410 TyrTyrIleSerAsnGlnTyrLysGlnMetProGlnGlyArgAsnLeuTyrLysIleGln 429
 QY 1604 TATGAGCGCGCGCGAGATGTACGCTCACACGCGCGGCTTCCCATAGCTGCTCC 1663
 Db 430 LeuThrAspHisThrAsnValLysCysLeu-----SerCysAsp 442
 QY 1664 ATGAGCCAG-----AATTCGACATGTTGTCAGCCACTAC 1699
 Db 443 LeuAsnProGlnArgCysGlnTyrTyrAlaValSerPheSerLysGlnAlaLysTyrTyr 462
 QY 1700 AGCAGCGTGAACAGCGCGCGCTGCGTGCACGCTTACAGATGAGCGCGCGCCGACAGCAGC 1759
 Db 463 GlnLeuGlyCysTrpGlnProGlyLeuProLeuTyrThrLeuHisArgSerThrAsp--- 481
 QY 1760 CCCCTGCACAGACAGCGCGCTTGGGCTAGCATGATGAGGCGACCGCTGC----- 1813
 Db 482 -----HisLysGlnLeuArg-----ValLeuGlnAspAsnSerAlaLeuAsp 495
 QY 1814 -----CCCGCGATATGTTCTCCAGAGATCTTCATTCCTCCACAGCGCGGCGAT 1864
 Db 496 ArgMetLeuGlnAspValGlnMetProSerLysLysLeuAspPheIleValLeuAsnGln 515
 QY 1865 GTGCGGCTC---TACGGCATGATCTACAGCCCGCTTGCAGCGCAGGAGAGAGCAGC 1921
 Db 516 ThrArgPheTrpTyrGlnMetIleLeuProHis---PheAspLysSerLysLysTyr 534
 QY 1922 CCCACCGCTCTTGTATATGAGAGCGCCCAAGTGCAGCTGGTGATATCTCTTCAA 1981
 Db 535 ProLeuLeuAspValTyrAlaGlyProCysSerGlnLysAlaAspAlaSerPhe--- 553
 QY 1982 GGCATCAAGTACTGGCGCTCAAC-----ACACTGGCGCTCCGCTGGCTAGCGCGCTG 2032
 Db 554 -----ArgLeuAsnTrpAlaThrTyrIleAlaSerThrGlnAsnIleIle 568
 QY 2033 GTTGTG---ATTGACGCGAGGCGCTCTGTACGCGAGGCTTGTGCGAAGGCGCGCTG 2089
 Db 569 ValAlaSerPheAspGlyArgGlySerGlyTyrGlnGlyAspLysIleMetHisAlaIle 588
 QY 2090 AAAAACCAATGGCGCAGGTGAGATCGAGACACAGGTGGAGGCGCTG---CAGTCTGTG 2146
 Db 589 AsnArgArgLeuGlyThrLeuGlnValGlnAspGlnIleGlnAlaAlaArgGlnPheVal 608
 QY 2147 GCGGAGATGATGGCTTCATGAGCCTGAGCGGAGTGTGCATGAGCTGGGCTGACGCG 2206
 Db 609 -----LysMetGlyPheValAspSerLysAspValAlaIleTrpGlnTyrSerTyrGly 626
 QY 2207 GCGTCTCTGCTATGAGGGGCTATTCACAAAGCCCGAGTGTCCATGAGTGGCCATCGCG 2266
 Db 627 GlyTyrValThrSerMetValLeuGlySerGlyValPheLysCysGlyIleAla 646
 QY 2267 GGTGCGCGGCTACCGCTGTGATGCGCTACGACAGAGGCTACGTAGCCTTACATG--- 2323
 Db 647 ValAlaProValSerArgTyrGlnTyrIleSerValTyrThrGlnArgTyrMetGly 666
 QY 2324 ---GACGTCCGAGACAAACAGCAGCAGCGATGAGCGGGGTGCGCGCTGACAGCTG 2380
 Db 667 LeuProIleProGlnAspAsnLeuAspHisIleTyrArgAsnSerThrValMetSerArgAla 686
 QY 2381 GAGAACTGCCAATGAGCCCAACCGCTGTATCTCTCCACGCTTCTGTGACGAGAAAC 2440
 Db 687 GlnHisPhe-----LysGlnValGlnTyrLeuLeuIleHisGlyThrAlaAspAspAsn 704
 QY 2441 GTGCACTTTTCCACACAAACTCTCTGCTCCCACTGATGCCAGCGGAGAAACTTAC 2500
 Db 705 ValHisPheGlnGlnSerAlaGlnIleSerLysAlaLeuValAspAlaGlyValAspPhe 724
 QY 2501 CAGCTCCAGATCTACCCCAAGAGAGACAGATATTCCTGCGCGCGAGTGGGAGAGAC 2560

```

DB 725 GlnAlaMetPrpTyrThrAspGluAspHisGlyIleAlaSerSerThrAlaHisGlnHis 744
OY 2561 TATGAGTCACTGGCTTCCTGACAGAA 2593
DB 745 IleTyrSerHisMetSerHisPheLeuGlnGln 755

RESULT 6
DPP6_HUMAN STANDARD; PRT; 865 AA.
AC P42658;
ID DPP6_HUMAN
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV like protein (Dipeptidyl aminopeptidase-
DE related protein) (Dipeptidylpeptidase VI) (DPPX).
GN DPP6
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=hippocampus;
RX MEDLINE=93372805; PubMed=8103397;
RA Yokotani N., Doi K., Wenthold R.J., Wada K.;
RT "Non-conservation of a catalytic residue in a dipeptidyl
RT aminopeptidase IV-related protein encoded by a gene on human
RT chromosome 7."
RL Hum. Mol. Genet. 2:1037-1039(1993).
CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, DPPX-L (SHOWN HERE) AND DPPX-S;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to jlicense@isb-sib.ch).
CC -----
DR EMBL: M96859; AAA35760.1; -
DR EMBL: M96860; AAA35761.1; -
DR MEROPS: S09.973; -
DR Gene: HGNC:3010; DPP6.
DR MIM: 126141; -
DR InterPro: IPR002469; DPPV_N_term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00326; Peptidase_S9; 1.
DR Pfam: PF00930; DPPV_N_term; 1.
KM Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT DOMAIN 1 95
FT TRANSMEM 1 116
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 117 865
FT CARBOHYD 173 173
FT CARBOHYD 319 319
FT CARBOHYD 404 404
FT CARBOHYD 471 471
FT CARBOHYD 535 535
FT CARBOHYD 566 566
FT CARBOHYD 813 813
FT VARSLIC 1 81

```

```

FT SQ SEQUENCE 865 AA: 97588 MW: 1481AE0E002464B CRC64;
Alignment Scores:
Pred. No.: 1,17e-15
Score: 452.00
Percent Similarity: 39.96%
Best Local Similarity: 23.87%
Query Match: 9.43%
DB: 1 Gaps: 43

US-09-976-674-4 (1-2617) x DPP6_HUMAN (1-865)
OY 76 CCGCTTCAGAGTCAGAGAACTGCTGGAGCGGCTCCGAGCATCAT---CCACGGCAG 132
DB 21 ProProGluAlaSerHisLeuGlnGlyGlnGlyProGluGlnAspGlyGlnGly 40
OY 133 CCGCACTACTCGGCGCTCATTTGCAACAAGCGCCCGACACTTCAGTTTGACAGAA 192
DB 41 AlaLysProLeuGlyProAlaGlnAlaAlaLysProArg-----Glu 55
OY 193 GACGAGTGAAGTCGGGCGCCACTCC-----CACCGCTCTACTGCGGAATGCC 243
DB 56 ArgGly-GlyGlyGlyGlyGlyAlaGlyGlyArgProArgPheGlnIleGlyArgse 75
OY 244 ATATGGCAGCGGAGAGAACTCCCTCTACTGAGATGCCAAGAGGTCGGAAGA 303
DB 75 RasPglYAspGlnGlnAspGlnGlnGlyGlnGlyGlnGlySerAsnProGlnArgsnTrpLysG 95
OY 304 GCGCTGCTGCTCTCTCTGTCGGAAGCAAGATGTCGATCATTTCCAGGCCACGCCACCA 363
DB 95 yIleAlaIleAlaLeuValIleLeuValIleCysSerLeuIleValIleThrSerValI 115
OY 364 TGGGCTCTACTCTCGGAGAGAGAGAGCTGCGAGGAGCGGAAACGCTGGGGCTTCGG 423
DB 115 eLeuLeuThrProAlaGlnAspAsnSerLeuSerGlnLysLysValIleThrValGluAs 135
OY 424 CATCACCCTCTACGACTTC-----CACAGCAGAGTGGCTCTCTCTTCGAGGCCAG 477
DB 135 pLeuPheSerGluAspPheLysIleHisAspProGluAlaLysTrpLe-----Se 152
OY 478 CAACAGCCTTTCACCTGTCGCGAC----- 502
DB 152 RasPTrpGluPheIleTyrArgGlnGlnLysGlyThrValArgLeuTrpAsnValGluTrh 172
OY 503 -----GCGCGCAAGAAC-----GCGCTT 519
DB 172 RasnThrSerThrValLeuIleGluGlyLysLysIleGluSerLeuArgAlaIleArgTy 192
OY 520 CATGGTGTCCCT-----ATGAAACCGCTGGAAT 549
DB 192 rGluIleSerProAspArgGlnGlyTrpAlaLeuPheSerTyrAsnValGluTrpLe---Ty 211
OY 550 CAAGACCACTGCTCAGGCGCCCGGATGACCCCAAAATGCTGCCCTGCGACTGCTT 609
DB 211 rGlnHisSerTyrTrpTyrValLeuSerLysIleProHisGlnGlyAspProGlnse 231
OY 610 CTTCCTCC-----TTCATCATTAACAGCGACCTGGGTGGCCAAATC---GAGACAG 660
DB 231 rIleuAspProGluGlnValSerAsnAlaLysLeuGlnTyrAlaGlyTrpGlyProLysG 251
OY 661 CGAG-----GAGCGCGCGCTGACCTTCTGC---CACCAAGTTTATC 699
DB 251 yGlnGlnLeuIlePheIlePheLeuLysAsnIleTyrTyrCysAlaHisValGlyLysG 271
OY 700 C-----AATGCTCTGGAGTGAACCCCAAGTCGCG-----GCTGGCGCACTT 741
DB 271 nAlaIleArgValValSerThrGlnLysGlnGlyValIleTyrAsnGlyLeuSerAspTr 291
OY 742 CGTCATACAGAGAGAG---TTCACCGCTTCACTGGTATCGTGGTGGCCGACAGCCTC 798
DB 291 pIeuTyrGlnGluGlnIleLeuLysThrHisIleAlaHisTrpTrpSerProAspGlyTh 311

```

```

QY 799 CTGGAGAGTTCAGAGGCGCTCAAGACGTCGCAATCCTGATAGAGATCGATGATC 858
   : |||: ||| :|||: |||
Db 311 r-----ArgleuAlaIleAlaIleAsnAspSe 321
QY 859 CGAGGTGAGGTGATTCAGTCCCTCT-----CTCGCGCTAGAGA 900
   : ||| :|||: ||| :|||: |||
Db 321 rArgValProIleMetCylLeuProThrIleGlySerIleTyrProThrValys-- 340
QY 901 AAGGAAGACGACTCGTATGGTACCCAGAGACAGCAAGATCCCAAGATGGCTT 960
   : ||| :|||: ||| :|||: |||
Db 341 -----ProIleHisTyrProIleValGlySerGluAsnProSerIleSerIle 356
QY 961 GAAACTGGTGAAGTCCAGACTGACAGCCAGCAAGATGCTCCGACCCAGAGAGAGA 1020
   : ||| :|||: ||| :|||: |||
Db 356 uHisValIle-----GlyLeuAsnGlyProThrHisAspLeuG1 369
QY 1021 GCTGGTGAACCCCTTCACTCGCTGTCCGAAAGTGGAG-----TACATCGCGAGGC 1074
   : |||: ||| :|||: ||| :|||: |||
Db 369 uMetMetProProAspAsp-----ProArgMetArgGluTyrTyrIleThrMetVa 386
QY 1075 CGGGTGGACCCGGGATGGCAATACGCTGGCCATGCTTCTGGACCGCGCCAGCACTG 1134
   : ||| :|||: ||| :|||: |||
Db 386 lLysTyrAlaIleThrSerThrIleValAla--ValThrTyrLeuAsnArgAlaGlnAsnVa 405
QY 1135 GCTCCAGCTGCTCTCTCCCGCGCGCTGTTCATCCGAGCAGACAGATGAGAGAGA 1194
   : ||| :|||: ||| :|||: |||
Db 405 lSerIleLeuThrLeuCysAspAlaIleThrIleGlyValCysThrIleIleGlySerHisGluAspG1 425
QY 1195 GCGGCTAGACCTCTGCCAGAGCTGTCCCGAGAAATGTCACCCGATGTGGTGTACAGAGA 1254
   : ||| :|||: ||| :|||: |||
Db 425 u----- 425
QY 1255 GGTCCACCAAGTGTGATC-----AATGTTATGATCATCTTTCATCCCTTCCCCCA 1305
   : |||: ||| :|||: ||| :|||: |||
Db 426 ----SerGluAlaIleTyrLeuHisArgGlnAsnGluIleProValPhe----- 439
QY 1306 ATCAGAGGAGAGAGAGAGTGTGCTTTCGCGCGCAATGAATGAAGAAGCGGCTCTG 1365
   : |||: ||| :|||: ||| :|||: |||
Db 440 -SerLysAspGlyArgIleGlySerPhePheIleArgAlaIlePro--GlnGlyGlyArgG1 458
QY 1366 CCATTGTGTACAAATGATCAGCCCGCTTTTAAATCCAGAGGCTAGCATTTGAGTGAGCCCTT 1425
   : ||| :|||: ||| :|||: |||
Db 458 yLysPheTyrHisIleThrVal-----SerSerSerGlnProAs 471
QY 1426 CAGCCCGGGGGAAGATGAATTAATGAGCCCATTAAGGAAGATTCCTGTACACACGG 1485
   : ||| :|||: ||| :|||: |||
Db 471 nSerSerAsnAspAsn-----IleGln-----SerIleThrSerG1 483
QY 1486 TGAATGGAGGTTTGGCGAGGACAGCGCTCCAGAGATCTGGGTC---AATGAGGAGACCA 1542
   : |||: ||| :|||: ||| :|||: |||
Db 483 yAspTTPAspVal-----ThrLysIleLeuAlaIleTyrAspGluIleGlyAs 498
QY 1543 GCTGGTGTACTTCCAGGAGCAGACAGACAGCGCTGAGACACACCTTACGTGTGAC 1602
   : |||: ||| :|||: ||| :|||: |||
Db 498 nLysIleTyrPheLeuSerThrGluAspLeuProArgArgIleGluIleuTyrSerAlaAs 518
QY 1603 CTATGAGGCGCGCGGAGATCTGACGCTCACCAGCCCGGCTTCCCAT----- 1654
   : |||: ||| :|||: ||| :|||: |||
Db 518 nThrGluGly-----AsnPheAsnArgGlnCysLe 528
QY 1655 -AGCTGTGCATGAGCCAGAACTTCGATGTTGCTGAGCCAGCAGACAGCGCTGAGACAC 1713
   : |||: ||| :|||: ||| :|||: |||
Db 528 uSerCysAspLeuValGlnCysThrTyrPheSerAlaSerPheSerHisSerMetAs 548
QY 1714 GCGCGCCCTGGCTGACGTCTACAGCTGACGCGCCCGAC-----GA 1755
   : |||: ||| :|||: ||| :|||: |||
Db 548 p-----PhePheLeuLeuIleuGlySerGluGlyProGlyValProMetValThrValHi 565
QY 1756 CGACCCCTCTCAGACAGACCGCCGCTTCTGGGCTAGCATGATGAG---GCAGCCACCTG 1812
   : |||: ||| :|||: ||| :|||: |||
Db 565 sasThrThrAspLysLysLysMetPheAspLeuGluThrAsnGluHisValLysLysAl 585
QY 1813 CCCCCCGGATTATGTTCTCTCCAGAGATCTTCATTTCCACACGCGC---TCGGATGTGGC 1869

```

```

Db 585 aIleAsnAspArgGlnMetProLysValGluTyrArgAspIleGluIleAspSerLys 605
QY 1870 GCTCTACGAGATGATTCACAAAGCCGCTTGCAGCCAGGGAAGAGACCCGCT 1929
   : ||| :|||: ||| :|||: |||
Db 605 nLeuProMetGlnIleLeuLysProAlaIleThrPheThrAspThrHisIleProIleuLe 625
QY 1930 CCTTTTGTATATGAGAGCCCGCCAGGTGCAAGCTGTGTATATACCTCTTCAAGGATCA 1989
   : ||| :|||: ||| :|||: |||
Db 625 uLeuValAlaAspGlyThrProGlySerGlnSerValAlaGluLysPheGlu--ValISe 644
QY 1990 GTACTTCGCGCTCAACACACATCTGCTTCCGCTAGCCGCTAGCCGCTGTGATTCAGCAG 2049
   : |||: ||| :|||: ||| :|||: |||
Db 644 rTyrGluThrVal---MetValSerSerHisGlyAlaValAlaValLysCysAspGlyArg 663
QY 2050 GGGCTCTGTACGAGAGGCTTCCGTTTCAAGGGGCGCTGAAAAACCAATGGGCGAGT 2109
   : |||: ||| :|||: ||| :|||: |||
Db 663 gGlySerGlyPheGlnGlyThrIleGlySerValLeuHisGluValAlaArgArgArgLeuLe 683
QY 2110 GGAATGAGAGACAGAGTGGAGGCGCTGCAAGTTCGTCGGCGAAGATAGGCTTCATGCA 2169
   : ||| :|||: ||| :|||: |||
Db 683 uGluGluLysAspGlnMetGluAlaValAlaArgThrMetLeu--LysGluGluTyrIleAs 702
QY 2170 CCTGAGCCGAGTTGCTTCCATGCTGCTGCTTACGAGGCGCTTCTCTGCTCATGAGGAGT 2229
   : |||: ||| :|||: ||| :|||: |||
Db 702 rArgThrArgValAlaValAlaPheGlyLysAspTyrGlyIleuThrIleuSerThrIle 722
QY 2230 AATCCACAG-----CCCAAGTGTTCAGAGTGGCCATCCGCGGCTGCCGCT 2277
   : ||| :|||: ||| :|||: |||
Db 722 uProAlaLysGlyLysLysGlnGluIleThrPheThrCysGlySerAlaLeuSerProI1 742
QY 2278 CACCGTGGATGGCCCTACACACAGAGGTACACGAGCTGATGAGATGCTCCGTGAA 2337
   : ||| :|||: ||| :|||: |||
Db 742 eThrAspPheLysLeuTyrAlaSerAlaPheSerGluArgTyrLeuLysLeuHisGlyLe 762
QY 2338 CAACACACAGCGCTATATGAGGCGGCTTCCGCGCGCTGACGTGAGAGAGCGCCATGA 2397
   : ||| :|||: ||| :|||: |||
Db 762 uAspAsnArgAlaIleTyrIleuMetThrLysValAlaHisArgValSerAlaLeu-----G1 780
QY 2398 GCGCCACCGCTTGTATCTTCCATCCAGCGCTTCCGAGGAGAAACGTGACATTTTCCAC 2457
   : ||| :|||: ||| :|||: |||
Db 780 uGluGlnGlnPheLeuIleIleHisProThrAlaAspGluLysIleHisPheGlnHisTh 800
QY 2458 AACTCTCTGTCTCCCACTGATCCGAGCAGGAGAAACCTTACAGCTCCAGATCTACCC 2517
   : ||| :|||: ||| :|||: |||
Db 800 rAlaGluLeuIleThrGlnLeuIleArgGlyLysAlaAsnTyrSerLeuGlnIleTyrPr 820
QY 2518 CAACGAGAGACAGATGTCGCTGCGCCGAGTCCGAGGAGCAGACATGACATGACGTTGCT 2577
   : |||: ||| :|||: ||| :|||: |||
Db 820 oAspGlnSerHisTyrPheThrSerSerLeuLysGlnHisLeuTyrArgSerIleI1 840
QY 2578 GCACCTTCTACAGGAA 2593
   : |||: ||| :|||: ||| :|||: |||
Db 840 eAsnPhePheValGlu 845

```

RESULT 7
DPP6_RAT
ID DPP6_RAT STANDARD: PRT; 859 AA.
AC P46101;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DiPeptidyl peptidase IV like protein (DiPeptidyl aminopeptidase-
DE related protein) (DiPeptidylpeptidase VI) (DPPX).
GN DPP6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92108018; PubMed=1729689;


```

Db 468 AsnAspAsn-----IleGln-----SerIleThrSerGlyAspIrr 479
QY 1493 GAGSTTTGGGAGGAGCGCTCCAGATCGGTGATGAGGAGACCACTGGTGTAC 1552
Db 480 AspValThr-----GluIleuThrTyrAspGluLysArgAsnLysLeuTyr 495
QY 1553 TTCAGGAGCAGGAGCAGCAGCGCTGGAGACCACTGTCGTGTCATGAGCGC 1612
Db 496 PheLeuSerThrGluAspLeuProArgArgArgHisLeuTyrSerAlaAsn----- 512
QY 1613 GCCGCGAGATCGTACGCGCTCCAGCGCCGCTTCCTCCATAGCTGC----- 1660
Db 513 -----ThrValAspAspPheAsnArgGlnCysLeuSerCysAsp 525
QY 1661 -----TTCATGAGCCGAGCACTTCGACATGTCGTC 1690
Db 526 LeuValGluAsnCysThrTyrValSerAlaSerPheSerHisAsnMetAspPheLeu 545
QY 1691 AGCCACTACAGCAGCGTGGAGCGCCGCC-----TGCCTGCAC----- 1729
Db 546 LeuLysCysGluGlyProGlyValAlaProThrValThrValHisAsnThrThrAspLysArg 565
QY 1730 ---GTCACAGCTGAGCGGCGCCGAGCAGCAGCCCTGCACAGCAGCAGCCGCTTCG 1786
Db 566 ArgMetPheAspLeuGluAlaAsnGlu-----GluValGlnLysAlaIleTyr 581
QY 1787 GCTAGCATGATGAGCGAGCGCAGCTGCCCGCATTTATGTTCTCCAGACATCTTCAT 1846
Db 582 AspArgGlnMetProLysIle-----GluTyrArgLysIleGluVal----- 595
QY 1847 TTCACAGCGCGCTGGATGTCGCTTACGCGATGATGATGATGATGATGATGATGATGAT 1906
Db 596 -----GluAspTyrSerLeuProMetGlnIleLeuLysProAlaThrPheThr 611
QY 1907 CCAGGAGAGAGCAGCAGCCCGCTCTTGTATATAGAGCGCCCGCAGCTGAGTGTG 1966
Db 612 AspThrAlaHisTyrProLeuLeuValAlaAspGlyThrProGlySerGlnSerVal 631
QY 1967 AATATACCTTCATCAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 2020
Db 632 SerGluArgPheGlu-----ValThrTrpGluThrValLeuValSerSerHis 647
QY 2021 GGCTAGCGCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2080
Db 648 GlyAlaValValValLysCysAspLysArgLysGlySerGlyPheGlnGlyThrLysLeuLeu 667
QY 2081 GGGGCGCTGAAAACCAATAGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2140
Db 668 HisGluValArgArgArgLeuGluPheLeuGlnGluLysAspGlnMetGluAlaValArg 687
QY 2141 TTCGTGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2200
Db 688 ThrMetLeu---LysGluGlnTyrIleAspLysThrArgValAlaValPheGlyLysAsp 706
QY 2201 TACGGGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2248
Db 707 TyrGlyGlyTyrLeuSerThrTyrIleLeuProAlaLysGlyLysGlnGlyIleThr 726
QY 2249 TTCAGAGTGGCCATCGCGGCTGCCCGGCTGACCGCTGATGAGCGCTGACGAGGATAC 2308
Db 727 PheThrCysGlySerAlaLeuSerProIleThrAspPheLysLeuTyrAlaSerAlaPhe 746
QY 2309 ACTAGAGCCTTCAATGAGCGTCCCTGAGAACACAGCAGCGCTATGAGCGGCTTCGCTG 2368
Db 747 SerGluArgTyrLeuGlyLeuHisGlyLeuAspAsnArgAlaTyrGlnMetThrLysLeu 766
QY 2369 GCGCTGACGCTGAGAGAGCTGCGCAATGAGCGCAACCGCTTGCATCTCCAGCGCTTC 2428
Db 767 AlaHisArgValSerAlaLeu-----GluAspGlnGlnPheLeuIleHisAlaThr 784
QY 2429 CTGAGCAGAAACGTGCACTTTTCCACAAACTCTCTGCTCCCACTGATGCGAGCA 2488
Db 785 AlaAspGluLysIleHisPheGlnHisThrAlaGluLeuIleThrGlnLeuIleLysGly 804

```

```

QY 2489 GGGAACTTACCAAGCTCCAGATCTACCCCAAGAGACACAGATTCGCTGCCCGAG 2548
Db 805 LysAlaSerTyrSerLeuGlnIleTyrProAspGluSerHisTyrPheHisSerValAla 824
QY 2549 TCGGCGGAGCAGCTATGAGTCACTGTCGTCGATCTTTCACAGGAA 2593
Db 825 LeuLysGlnHisLeuTyrArgSerIleIleGlyPhePheValGlu 839

RESULT 8
ID DP66_BOVIN STANDARD; PRT; 863 AA.
AC P42659;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV like protein (dipeptidyl aminopeptidase-
related protein) (dipeptidyl peptidase VI) (DPPX).
GN DPP6.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX Bovidae; Bovinae; Bos.
NX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=92108018; PubMed=1729689;
RA Wada K., Yokotani N., Hunter C., Doi K., Wentholt R.J., Shimasaki S.;
RT "Differential expression of two distinct forms of mRNA encoding
RT members of a dipeptidyl aminopeptidase family.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: DPPX-L (SHOWN HERE) AND DPPX-S;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
CC BRAIN, KIDNEY, OVARY AND TESTIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M76428; AAC41622.1; -.
DR EMBL; M76429; AAC41623.1; -.
DR MEROPS; S09_973; -.
DR InterPro; IPR002469; DPPIV_N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_sstr_site.
DR Pfam; PF00930; Peptidase_S9; 1.
DR Pfam; PF00930; DPPIV_N term; 1.
KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT DOMAIN 1 93
FT TRANSDOM 94 114
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT EXTRACELLULAR (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 171 171
FT CARBOHYD 402 402
FT CARBOHYD 469 469
FT CARBOHYD 533 533
FT CARBOHYD 564 564
FT CARBOHYD 811 811
FT VARSPLIC 1 79
FT MASLORTKINSTRSPAPPEASRLIGGGPEEDAGPK
FT PLGAQPAARAREGGGAGGRFRFYQARSDDDED ->

```



```

Db 775 SerAlaLeu-----GluGlyGlnGlnPheLeuValIleHisAlaThrAlaAspGluLys 792
QY 2441 GTGACATTTTCACACAAACTTCCTCCTCCCAACGATGCCAGCGAGGAACCTTAC 2500
Db 793 IleHisPheGlnHisThrAlaGlnLeuIleThrGlnLeuIleGlyLysAlaAsnIyr 812
QY 2501 CAGCTCCAGATCTACCCCAAGACAGACAGATTCCTGCTGCCGAGTCCGCGAGCAGC 2560
Db 813 SerLeuGlnIleIyrTrpProAspGlnSerHisIyrPheSerAlaAlaLeuGlnGlnHis 832
QY 2561 TATGAACTGACGTTGCTGCTACCTTCTACAGGAA 2593
Db 833 LeuHisArgSerIleLeuGlyPhePheValGlu 843
RESULT 9
SEPR_MOUSE STANDARD: PRT: 761 AA.
ID SEPR_MOUSE
AC P97321;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
GN membrane serine protease).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX STRAIN-BALB/C; TISSUE-Embryo;
MEDLINE-97284459; Pubmed-9139873;
RA Niedermeyer J., Scanlan M.J., Garin-Chesa P., Dalber C., Fiebig H.H.,
RA Old L.J., Rettig W.J., Schnapp A.;
RT "Mouse fibroblast activation protein: molecular cloning, alternative
RT splicing and expression in the reactive stroma of epithelial
RT cancer."
RL Int. J. Cancer 71:383-389(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE-Breast;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and contribute to invasiveness in malignant
CC cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive (by similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (by similarity).
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms: 1 (shown here), 2 and 3; are
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
CC embryos from day 7-19 and in new-born mice (P1).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y10007; CAAT71116.1; -.
DR EMBL: BC019190; AAH19190.1; -.
DR MEROPS: S09.007; -.
DR MGD: MGI:109608; FAP.
DR InterPro: IPR002469; DPPIV_N_term.
DR InterPro: IPR001375; Peptidase_S9.

```

```

DR InterPro: IPR002471; Prol_endopep_ser.
DR InterPro: IPR000379; Ser_estr_site.
DR Pfam: PF00326; Peptidase_S9; 1.
DR Pfam: PF00930; DPPIV_N_term; 1.
DR ProSite: PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT ACT_SITE 26 761 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 624 624 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 702 702 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 734 734 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 31 35 MISSING (IN ISOFORM 2).
FT VARSPPLIC 31 63 MISSING (IN ISOFORM 3).
FT CONFLICT 737 737 S->L (IN REF. 2).
SQ SEQUENCE 761 AA; 87944 MW; 9174C3AEDA213B25 CRC64;

```

Alignment Scores:

```

Pred. No.: 5,41e-15 Length: 761
Score: 438.50 Matches: 183
Percent Similarity: 37.45% Conservative: 126
Best Local Similarity: 22.18% Mismatches: 287
Query Match: 9.14% Indels: 229
DB: Gaps: 35

```

US-09-976-674-4 (1-2617) x SEPR_MOUSE (1-761)

```

QY 362 CATGGGCTCTCTCTCTCGAGAGAGAGCTGTCGAGGAGCGGAACGCTGGGGTCTTC 421
Db 79 TyrAsnIleGlnThrArgGlnSerIyrIleIleLeuSerAsnSerThermIysSerVal 98
QY 422 GGCATCACCTCTCTACGACTTC-----CAC 445
Db 99 AsnAlaThrAspIyrGlyLeuSerProAspArgGlnPheValIyrLeuGlnSerAspIyr 118
QY 446 AGCGAGATGGGCTCTCTCTCTTCAGCCACGACCAACCTCTTCACTGTCGCGAGCG 505
Db 119 SerIysLeuTrpArgIyrSerIyrThrAlaThrIyrIleIyrAspLeuGlnAsnGly 138
QY 506 GCGAAGACGGCTTCATGCTGCTCCCTATGAACCGCTGGAATCAAGACCGAGCTCA 565
Db 139 -----GluPheValArgGlyIyrGlnLeuProArgProIleGlnIyrLeuGlyStrp 155
QY 566 GGGCCCGGATGACCCCAAAATCTGC-----592
Db 156 SerPro--ValGlySerIysLeuAlaIyrValIyrGlnAsnAsnIleIyrLeuLysGln 174
QY 593 ---CCTGCCGACCGCTCTCTCTCTTCATCAATACAGCGACTGTGGTGGCCAC 649
Db 175 ArgProGlyAspProPhe-----GlnIle 183
QY 650 ATCGACAGCGGAGCGGCGGCTGACCTTCTGCACCAAGTTTATCCAAATGCTCTG 709
Db 184 ThrIyrThrGlyArgGlnAsnArgIle-----192
QY 710 GATGACCCCAAGTGTGCGGTGTGGCCACTTCTGTCATACAGAGAGCTTGCAGCGCTTC 769
Db 193 -----PheAsnGlyIleProAspTrpValIyrGlnGlnGlnMet---LeuAla 207
QY 770 ACTGGGTAC-----TGGTGGTGCCGCCACAGCTCTCTGGGAAGTTTCAGAGGGCTCTAG 823
Db 208 ThrIyrIyrAlaLeuIyrTrpSerProAspGlyLysPhe-----220
QY 824 ACGTGCAGATCTCTATAGAGAGAGTCAATGATGATCGGAGTGTGAGGTGATTT-----874

```

Db 221 -----LeuAlaTyrValGluPheAsnAspSerAspIleProIleIleAlaTyrSer 237
 QY 875 -----CAGCTCCCTCTCTGCGGTAGAGAAGAGAGAGAGCTGAT 919
 Db 238 TyrTyrGlyAspGlyGlnTyrPro-----ArgThrIleAsnIle 250
 QY 920 CCGTACCCAGAGACAGACAGAGAAATCCAGATGCCCTTGAAATGCGTAGCTCCAG 979
 Db 251 ProTyrProIysAlaGlyAlaIysAsnProValValArgVal----- 264
 QY 980 ACTGACAGCCAGGCGAGATGCTGACACCAGAGAAAGAGAGCTGTCAGCCCTTCAGC 1039
 Db 265 -----PheIleValAspThrThrTyrProHisValIleGlyProMetGlu 279
 QY 1040 TCGGTGTCCCGAGAGTGAATACATGCGAGCGCGGTGAGCCGGATGCCAAATAC 1099
 Db 280 -----ValProValIleProGluMetIleAla-----SerSerAspTyrTyrPhe 293
 QY 1100 GCCTGGGCGCATGTCCTGAGACCGGCCAG-----CAGTGGCTCCAGCTGCTCTC 1150
 Db 294 SerTyrPheThrTyrValSerSerGluArgValCysLeuGlnTyrLeuIysArgValGln 313
 QY 1151 CTCGCCCGGCGCTGATCCGACGACAGAGATAGAGAGAGCGGCTAGCCTCTGCC 1210
 Db 314 AsnValSerValLeuSerTleCysAspPheArgGluAspTrpHis-----Ala 329
 QY 1211 AGAGCTGTCCCGAGAGATGTCAGCCGTATGCTGACGAGAGAGTACCAGCATCTGG 1270
 Db 330 TyrPheIysProIysAsn-----GlnGluHisVal-----GluGluSerArgThrGlyTyr 346
 QY 1271 ATCATTTATGACATCTTCTATCCCTTCCCGCAATCAGAGAGAGAGAGAGCTCTGC 1330
 Db 347 AlaGlyGly-----PhePheValSerThrProAlaPheSerGlnAspAlaThrSer 363
 QY 1331 TTTCTCGCGCAATGATGACAGACCGGCTTGCATTTGTATCAAGTACCGCGCTT 1390
 Db 364 TyrTyrIysIlePheSerAspIysAspGlyTyrIysHisIleHisTyrIle----- 380
 QY 1391 TTAAATCCAGGCGTACGATGTGAGTGAAGCCCTTCAGCCCGGGAGAAATTTAG 1450
 Db 381 -----Lys 381
 QY 1451 TGCCCATTTAAGAGAGATGCTCTGACACGCGTGAATGGAGTTTGGCCAGGCAC 1510
 Db 382 AspThrValIleAsnAlaIleGlnIleThrSerGlyIysTrpGluAlaIle----- 398
 QY 1511 GGCTCCAAAGTGGTGCATGAGAGACCAAGCTGGTAC-----TTCCAG 1558
 Db 399 -----TyrIlePheArgValThrGlnAspSerLeuPheTyrSerAsnGluPheGlu 416
 QY 1559 GGCACCAAGACACGCGCTGGAGCACACCTTACGCTGCTATGAGCGCGCGC 1618
 Db 417 Gly-----TyrProGlyArgArgAsnIleTyrArgIleSerIleGlyAsnSerPro 433
 QY 1619 GAATGCTAGGCTCACACCGCGCGCTTCCCATAGCTGCTCCATG----- 1666
 Db 434 ProSerIysLysCysValThr-----CysHisLeuArgGlyGluArg 447
 QY 1667 -----AGCCGAACTTCGACATGTTGCTGACGACCTACGACAGAGAGCAGC 1714
 Db 448 CysGlnTyrTyrThrAlaSerPheSerTyrLysAlaIysTyrTyrAlaLeuValCysTyr 467
 QY 1715 CCGCCCTGCGTGCAGCTCTCAACCTG--AGCGCGCCGACGACGACCCCTGAC-- 1768
 Db 468 GlyProGlyLeuProIleSerThrLeuHisAspGlyArgThrArgGlnGluVal 487
 QY 1769 -----AAGCAGCCCGCG----- 1780
 Db 488 LeuGluGluAsnIysGluLeuGluAsnSerLeuArgAsnIleGlnLeuProIysValGlu 507
 QY 1781 -----TTCTGGGCTAGCATGATGAGGAGCGACG 1810
 Db 508 IleLysLysLeuLysAspGlyGlyLeuThrPheThrTyrIysMetIle----- 523

QY 1811 TGCCCCCGGATTTATTTCTCCAGAGATCTTCCATTTCCACAGCGCGCTGGATGCGG 1870
 Db 524 -----LeuProProGln----- 527
 QY 1871 CTCTAGGCGATGATCTACAAAGCCCGCTTGACGCCAGGGAAGAACGCCACCGCTC 1930
 Db 528 -----PheAspArgSerIysLysTyrLeuLeu 537
 QY 1931 CTCTTTGATATGAGGCGCCCGACAGCTGACCTGCTGAATACCTCTTC-----AAAGCC 1994
 Db 538 IleGlnValTyrGlyGlyProCysSerGlnSerValIysSerValPheAlaValAsnTyr 557
 QY 1985 ATCAAGTACTTGCGGCTCAACACACTGCGCTCCCTGGGCTACGCGGCTGTGATGAC 2044
 Db 558 IleThrTyrIleu-----AlaSerLysGluIleValIleAlaLeuValAsp 573
 QY 2045 GGCAGGGGCTCTGTCAGCGAGGCGTTCGCTCGAGAGGCGCTGAAAACCAATAGGC 2104
 Db 574 GlyArgGlyThrAlaPheGlnGlyAspLysPheLeuHisAlaValTyrArgLysLeuGly 593
 QY 2105 CAGGTGAGATCCGAGCACCGAGTGGAGGCGCTGCAGTTCTGGCGGAGATAGCTTC 2164
 Db 594 ValTyrGluValGluAspGlnLeuThrAlaValArgLysPheIleGlu--MetGlyPhe 612
 QY 2165 ATGACCTGAGCGCGAGTGCATTCATGCTGCTGCTACGAGGCGCTCTCTGCTCATG 2224
 Db 613 IleAspGluArgIleAlaIleTyrPheTyrSerTyrGlyGlyTyrValSerSerLeu 632
 QY 2225 GGGCTAATCCAAAGCCCGAGGCTTTCAGAGTGGCCATCGCGGCTGCCCGGTACCGCTC 2284
 Db 633 AlaLeuAlaSerGlyThrGlyLeuPheLysCysGlyIleAlaValAlaProValSer 652
 QY 2285 TGGATGGCTACGACAGGCTACATGAGCGCTACATGAGCGCT-----GAGAAC 2338
 Db 653 TyrPheTyrTyrAlaSerIleTyrSerGluArgPheMetGlyLeuProThrLysAsp 672
 QY 2339 AACGAGCAGCGCTATGAGCGCGGCTCCGTGCGCTGACGCTGAGAAAGTCCCAATGAG 2398
 Db 673 AsnLeuGlnHisTyrLysAsnSerThrValMetAlaArgAlaGluTyrPheArgAsnVal 692
 QY 2399 CCCAAGCGCTTGCATTATTCCTCCAGCGCTCTCGACGCAAAACGTGCTTTTCCACACA 2458
 Db 693 Asp-----TyrLeuLeuIleHisGlyThrAlaAspAspAsnValHisPheGlnAsnSer 710
 QY 2459 AACTCTCTGCTCCCAAGTATGCTCCGAGCGAGCAAACTTACCGACTCCAGATTCACCC 2518
 Db 711 AlaGlnIleAlaLysAlaLeuValAsnAlaGlnValAspPheGlnAlaMetTyrTyrSer 730
 QY 2519 AACGAGACACAGTATTCGCTGCGCGCGAGTCCGCGAGCAGCTATGAACTCAGCTGCTG 2578
 Db 731 AspIleAsnHisGlyIleSerSerGlyArgSerGlnAsnHisLeuTyrThrHisMetThr 750
 QY 2579 CACTTCTACAGAA 2593
 Db 751 HisPheLeuLysGln 755
 RESULT 10
 DAP2_YEAST
 ID DAP2_YEAST STANDARD; PRT; 818 AA.
 AC P18962;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DAP B) (YSCV).
 GN DAP2 OR YHR028C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89174971; PubMed=2647766;

Accession	Protein Name	Length	Score	Similarity
RA	Robertson, C.J., Pohlig G., Rothman J.H., Stevens T.H.;	818	1.09e-14	39.47%
RT	"Structure, biosynthesis, and localization of dipeptidyl	164		
RT	aminopeptidase B, an integral membrane glycoprotein of the yeast			
RT	vacuole."			
RL	J. Cell Biol. 108:1363-1373(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-S288C / AB972;			
RX	JMEDLINE=94378003; PubMed=8091229;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,			
RA	Ku D., Favelli A., Fulton L., Gattung S., Geisel C., Kirsten J.,			
RA	Duena T., Hillier L., Jier M., Johnston L., Langston Y.,			
RA	Latrelle P., Louis E.J., Macri C., Maris E., Menezes S., Mouser L.,			
RA	Mhan M., Rifkin L., Riles L., St Peter H., Trevasis E., Vaughan K.,			
RA	Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,			
RA	Vaudin M.;			
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome			
RT	viii".			
RL	Science 265:2077-2082(1994).			
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE			
CC	VACUOLES.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X15484; CAA33512.1; -;			
DR	EMBL; U10399; AAB68879.1; -;			
DR	PIR; A30107; A30107.			
DR	PIR; S46780; S46780.			
DR	MEROPS; S09_006; -;			
DR	SGD; S0001070; DAP2.			
DR	InterPro; IPR002469; DDPV_N_term.			
DR	InterPro; IPR001375; Peptidase_S9.			
DR	InterPro; IPR002471; Prol_endopep_ser.			
DR	InterPro; IPR000379; Ser_estr_site.			
DR	Pfam; PF003326; Peptidase_S9_1.			
DR	Pfam; PF00930; DDPV_N_term_1.			
DR	PROSITE; PS00708; PRO_ENDOPEP_SER_1.			
DR	Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;			
RW	Transmembrane; Glycoprotein; Signal-anchor.			
FT	DOMAIN	1	29	
FT	TRANSMEM	30	45	
FT	DOMAIN	46	818	
FT	ACT_SITE	679	679	
FT	ACT_SITE	756	756	
FT	ACT_SITE	789	789	
FT	CARBOHYD	63	63	
FT	CARBOHYD	79	79	
FT	CARBOHYD	110	110	
FT	CARBOHYD	139	139	
FT	CARBOHYD	372	372	
FT	CARBOHYD	392	392	
FT	CARBOHYD	421	421	
FT	CARBOHYD	738	738	
FT	CONFLICT	83	83	
FT	CONFLICT	125	125	
FT	CONFLICT	182	188	
FT	CONFLICT	200	200	
FT	CONFLICT	366	375	
FT	CONFLICT	808	818	
SO	SEQUENCE	818 AA;	93404 MM;	318F450445375BD3 CRC64;

Best Local Similarity:	25.39%	Mismatches:	262
Query Match:	9.02%	Dbs:	129
	1	Gaps:	28

US-09-976-674-4 (1-2617) x DAF2_YEAST (1-818)

QY	728	GGTGTGGCGACACCTTCGTCATACAGGAAG---	TTGACCGGCTTCAGTGGTACTGGTG	784
Db	234	GlyIysProAspTrpValTYrGluGluGluValPheGluAspLysAlaIaIaTrpTr		253
QY	785	TGCCCCACAGCCTCTCCGGAAGGTGACAGAGGCGCTCAAGACCTCGATCTATGAG		844
Db	254	SerProThrGlyAspTYr		263
QY	845	GAAGTCGATGAGTCCGAGGTGAGGTCAATTCAGCTCCCTCTCCGCTAGAGAAG		904
Db	264	LysIleAspGluSerGluValGluPheIleIleProTYrValGluInsPrluLys		283
QY	905	-----AAGACGACCTGTATCGGTACCCCAAGACAGCAAGATCCCAAG		952
Db	284	AspIleTYrProGluMetArgSerIleLysTYrProLysSerLysIleProAsnProHis		303
QY	953	ATTCGCTTGAACACTGGCTGAGTTCGACACTGACAGCAGGCAAGATCGTCCGACCAG		1012
Db	304	AlaGluLeuTrpValTYrSerMetLysAspGly		314
QY	1013	GAGAAAGAGCTGTGACAGCCCTTCAGCTCCGCTCCCGAAGTGGAG---		1060
Db	315	-----ThrSerPheHisProArgIleSerGluLysInsLys		326
QY	1061	-----TTCATCCGACGAGGCGGGGTGGACCCGGATGGCAATACGCTGGGC		1106
Db	327	AspGlySerLeuLeuIleThrGluValThTrpValGlyAsnLysAn---		1345
QY	1109	ATGTCCTCGACCGCGCCCGACAGACTGGCTCCAGCTGCTCTCCCTCCCGGCGCTTC		1168
Db	346	LysThrTrpAspArgSerSerAspIleLeuThValPheLeuIleAspThr		362
QY	1169	ATCCCGAGACACAGAAATGAGAGCAGCGGCTAGCTCTCCGACAGCTGCCAGAAAT		1228
Db	363	IleAlaLysThrSerAn		368
QY	1229	GTCACGCGGTATGGTGIGTACAGAGAGGTACACAC---		1287
Db	369	-----ValValArgAsnGluSerSerAsnLysLysTYrTrpIleThrHis		384
QY	1283	GACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGAGAGAGCTGCTTCGCGAGCC		1344
Db	385	AsnThrLeu		391
QY	1343	AATGAAATGC---		1390
Db	392	AsnGluThrPheAspArgProHisAsnGlyTYr		405
QY	1391	TTAAATATCCACGGCTACGATGGAGTGAAGCCCTTC---		1447
Db	406	LeuProIleGlyGlyTYrAsnHisLeuAlaTYrPheGluAsnSerAnSerSerHisTYr		425
QY	1448	AAGTCCCATTAAGAAAGAAATGCTGCTCCACGAGCTGAAGGAGGAGGTGGGAGAG		1507
Db	426	Lys		436
QY	1508	CACGGCTCCAAATCTGGGTCAATGAGAGAGACCAAGCTGGTACTTCCAGGGCACCAAG		1567
Db	437	AsnGluProLeuAlaPheAspSerMetGluAsnArgLeu---		455
QY	1568	GACACGCGGCTGGAGACACCACTTACGTGGTCACTATGAGGCGGCGGAGATCGTA		1622
Db	456	LysSerSerThrGluArgHisValTYrTYrIleAspLeuArgSerProAsnGluIleIe		475
QY	1628	CGCCCTCACAGCGCCGCG---		1677
Db	476	GluValThrAspThrSerGluAspGlyValTYrAspValSerPheSerSerGlyArgG		495

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: U09278; AAB49652.1; -
 DR EMBL: U76833; AAC51668.1; -
 DR EMBL: AF007822; AAF21600.1; -
 DR MEROPS: S09.007; -
 DR Genew: HGNC:3590; FAP.
 DR MIM: 600403; -
 DR InterPro: IPR002469; DPIP_N-term.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; Pro_endopep_ser.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR Pfam: PF00930; DPIP_N-term; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SFR; 1.
 KM Hydroxylase; Protease; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Alternative splicing.
 FT DOMAIN 1 4
 FT TRANSMEM 5 25
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 26 760
 FT ACT_SITE 624 624
 FT ACT_SITE 702 702
 FT ACT_SITE 734 734
 FT CARBOHYD 49 49
 FT CARBOHYD 92 92
 FT CARBOHYD 99 99
 FT CARBOHYD 314 314
 FT CARBOHYD 679 679
 FT VARSPIC 1 521
 FT CONFLICT 207 207
 FT CONFLICT 229 229
 FT CONFLICT 354 354
 FT SEQUENCE 760 AA; 87820 MM; AOD34B801BE07EA CMC64;
 Alignment Scores:
 Pred. No.: 2.28e-14 Length: 760
 Score: 426.00 Matches: 168
 Percent Similarity: 38.85% Conservative: 109
 Best Local Similarity: 23.56% Mismatches: 256
 Query Match: 8.88% Indels: 180
 DB: 1 Gaps: 31

US-09-976-674-4 (1-2617) x SEPR_HUMAN (1-760)

OY 593 CCTGCGACCTGCTTCTCTTCATCAATAACAGCAGCAGCTGTGGTGGCCAAATC 652
 DB 176 ProglinspProprophe- 181
 OY 653 GAGACAGCGAGGCGGCGCTGACCTTCCACCAAGTTTATCCATGCTCGAT 712
 DB 182 -----GlnIleThrphe-----AsnGlyArgGluAsnIleIle 193
 OY 713 GACCCCAAGTGTGGGGTGGCCACTTGTCTATACAGGAAGATTGACCGCTTCACT 772
 DB 194 Asn-----GlyIleProAspTyrPValTyrGluGluGluMet---LeuProthr 208
 OY 773 GGGTAC-----TGGTGTGGCCCAACAGCCTCCGCGGAAGTTGACAGGGCCTCAAGAC 826
 DB 209 LysTyrAlaLeuTyrPThrPserProAsnGlyLysPhe----- 220
 OY 827 CTGGCACTCTATGAGAGAGATGATCCGAGGTGAGGTATTCACGTCCCTCT 886
 DB 221 -----LeuAlaTyrAlaGluPheAsnAspLysAspIleProValIleAlaTyrSerTyr 238
 OY 887 CCTGCGCTAGAGAA--AGGAGACGAGCTGATCGGTACCCAGAGACAGCAGCAAG 943
 DB 239 TyrGlyAspGluGlnTyrProArgThrIleAsnIleProTyrProLysAlaGlyAlaLys 258

OY 944 AATCCCAAGATTGCTTGAATACTGCTGAGTTCACAGCTGACAGCCAGGCAATCGTC 1003
 DB 259 AsnProValAlaArgIle-----PheIleIle 267
 OY 1004 TCGACCCAGAGAAAGACCTGTGCGAGCCCTTCACCTGCTGTCCGAAGTG----- 1057
 DB 268 AspThrThrTyrProAlaTyrAlaGlyProGluGluValProValProAlaMetIleAla 287
 OY 1058 -----GAGTACATGCGCAGGCGCGGGTGGACCCGGGATGCAAAATACGCTGGGCC 1108
 DB 288 SerSerAspTyrTyrPheSerThrPheThrValThrAspGluArgValCys----- 305
 OY 1109 ATGTTCCTGGACCGGCCCGACAGCGCTCCAGCTGCTCTCTCTCTCTCTCTCTCT 1168
 DB 306 -----LeuGlnIlePheLysArgValGlnAsnValSerValLeuSer 319
 OY 1169 ATC-----CCGACGACAGCAAT----- 1186
 DB 320 IleCysAspPheArgGluAspTyrPheThrThrPheAspCysProLysThrGlnIleHisIle 339
 OY 1187 GAGAGCAGCGGCTA-----GCTCTGCGCAGAGCTGTCCCGAGAATGTCACGCGTAT 1240
 DB 340 GluGluSerArgThrGlyTyrPheAlaGlyGlyPhePheValSerArgProValPheSerTyr 359
 OY 1241 GTGGTGTACGAGGAGGTGACCAACGCTGATCATGTCATCTTATCCCTTC 1300
 DB 359 ----- 359
 OY 1301 CCCCAATCAGAGGAGAGGAGCAGCTGCTTCTCCCGCCCAATGAATGCAAGACCGGC 1360
 DB 360 -----AspAlaIleSerTyrTyrLysIlePheSerAspLysAspGly 373
 OY 1361 TTCTGCCATTGTACAAAGTACCGCCGTTTAAATCCAGGCGTACGATTGAGTAG 1420
 DB 374 TyrLysHisIleHisTyrIle----- 380
 OY 1421 CCTCTACCGCCCGGGAAGATGATTTAAGTCCCATTAAGAGAGATTGCTGTGAC 1480
 DB 381 -----LysAspThrValGluAsnAlaIleGlnIleThr 391
 OY 1481 AGCGGTGATGAGAGGATTTTGGCGAGCAGCGCTCCAGATGTGCGTCAATGAGAGACC 1540
 DB 392 SerGlyLysTyrPheGluAlaIle-----AsnIlePheArgValThrGlnAsp 406
 OY 1541 AAGCTGTGTACTTCCAGGGCACC---AAGACACGCGCGTGGAGCACCACCTGTAC--- 1594
 DB 407 SerLeuPheTyrSerSerAsnGluPheGluGluTyrProGlyArgArgAsnIleTyrArg 426
 OY 1595 -----GTGGTCACTATGAGGCGCGCGGCGAGATCGTACGCTTACACGCGCGGCTTC 1648
 DB 427 IleSerIleGlySerTyrProProSerLysCysVal----- 439
 OY 1649 TCCCATACCTGCTCATG-----AGCCAGCAACTTCGACATG 1684
 DB 440 -----ThrCysHisLeuArgLysLuarGlyCysGlnTyrTyrThrAlaSerPheSerasp 457
 OY 1685 TTCTGACCCACTACAGCAGCGTGAAGCGCGCGCTCGTGCACGTTACAAAGCTG--- 1741
 DB 458 TyrAlaLysTyrTyrAlaLeuValLysTyrGlyProGlyIleProIleSerThrLeuHis 477
 OY 1742 AGCGGCCCGGAGAGACCCCTGAC-----AAGCAGCCCGGCTTCTGGCTGACATG 1795
 DB 478 AspGlyArgThrAspGluGluIleLysIleLeuGlnAsnLysGluLeuGluAsnAla 497
 OY 1796 ATGAGCAGCAGCAGCTGCCCGCGGATTAATGTTCTCCAGAGATTTCCATTTCCACACG 1855
 DB 498 LeuLysAsnIleGluLeuProLysGlu-----GluIleLysLysLeuGluVal 513
 OY 1856 CGCTCGATGTGCGGCTTACGCGCATGTACTAAGCCCAAGCCGCTTGCAGCCAGGAG 1915
 DB 514 AspGluIleThrLeuTyrTyrLysMetIleLeuProGln---PheAspArgSerLys 532

```

QY 1916 AAGCACCACCCCTCTCTTTGTAATGAGAGCCGCCAGGTGAGCTGGTAATACCTCC 1975
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 533 LysTrpProLeuLeuLeuValGlyTrpGlyProCysSerGlnValArgSerVal 552
QY 1976 TTC-----AAGGCATCAACTACTGGGGCTCAACACACTGGCGCTCCCTGGGCTACGCC 2029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 PheAlaValAsnTrpIleSerTyrLeu-----AlaSerLysGlnGlyMetVal 568
QY 2030 GTGGTGTGATGATGAGCGGGGGCTCTGTCAGCGAGGGCTTCGGTTCGAGAGGCCCTG 2089
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 569 IleAlaLeuValAspGlyArgGlyThrAlaPheGlnGlyAspLysLeuTyrAlaVal 588
QY 2090 AAAAACCAATGGCCAGGTGAGATGAGAGCAGGAGCGGCGCTGACGTGCTGGGCC 2149
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 589 TyrArgLysLeuGlyValIleGlyValGlnValGlnValGlnValAlaValArgLysPheIle 608
QY 2150 GAGAAAGATGGCTCATGACCTGAGCGGAGTTCACATCCATGGCTGGTCTACGGGGCC 2209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 Gln--MetGlyPheIleAspGlnLysArgIleAlaIleTrpGlyTyrPserTyrGlyGly 627
QY 2210 TTCCTCTGCTCATGGGCTTAATCCACAGCCCGAGGTGTCAAGTGGCCATGCGGGGT 2269
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 628 TyrAlaSerSerLeuAlaLeuAlaSerGlyThrGlyLeuPheLysCysGlyIleAlaVal 647
QY 2270 GCCCGGTCACCGTCTGATGGCTACGACAGGAGGTAACACTGAGCGCTACATGAGAGCTC 2329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 648 AlaProValSerSerTrpGlnTyrTyrAlaSerValTyrThrGlnArgPheMetGlyLeu 667
QY 2330 CCT-----GAGAACACACAGCAGCGCTATGAGCGGCTCCGCTGCGCTCAGCTGAG 2383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 ProThrLysAspAsnLeuGlnHisTyrLysAsnSerTrpAlaMetAlaArgAlaVal 687
QY 2384 AAGCTGCCAATGAGCCCAACCGCTGCTTATCCCAAGCTCTCTGCGAGAAACGTG 2443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 688 TyrPheArgAsnValAsp-----TyrLeuLeuIleHisGlyThrAlaAspAsnVal 705
QY 2444 CACTTTTCCACACAACTCTCTGCTCCCACTGATCCGAGAGGAGAACTTACCAG 2503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 HisPheGlnAsnSerAlaGlnIleAlaValAlaLeuValAsnAlaGlnValAspPheGln 725
QY 2504 CTCGAGTCTTATCCCAACGAGACAGACTTTCGCTGCCCGGAGCTGGCGC----- 2554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 AlaMetTrpTyrSerAspGlnAsnHisGlyLeu-----SerGlyLeuSerThr 741
QY 2555 GAGCAGTATGAACTGACCTGCTGCTCTTCTACAGGAA 2593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 742 AsnHisLeuTyrThrHisMetThrHisPheLeuLysGln 754

```

```

RL (In) Gething M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.):
RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RL Press, Oxford (1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96437977; PubMed=8840505;
RA Galisson F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RT chromosome XV of the yeast Saccharomyces cerevisiae.";
RL Yeast 12:877-885(1996).
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC ALPHA-FACTOR PRECURSOR.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, LYSOSOME-LIKE
CC VACUOLES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L21944; AAA35119.1; -
DR EMBL: U08230; AAA17897.1; -
DR EMBL: X92441; CAA63182.1; -
DR EMBL: Z75127; CAA99437.1; -
DR PIR: A49737; A49737.
DR PIR: S45451; S45451.
DR MEROPS: S09.005; -.
DR SGD: S0005745; STE13.
DR InterPro: IPR002469; DDPV N.term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; ProL_endopep_ser.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00326; Peptidase_S9; 1.
DR Pfam: PF00930; DDPV N.term; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; FALSE NEG.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor; Phorbome response.
KW DOMAIN 1 119
FT TRANSMEM 120 140
FT DOMAIN 141 931
FT ACT_SITE 785 785
FT ACT_SITE 863 863
FT ACT_SITE 896 896
FT CARBOHYD 377 377
FT CARBOHYD 814 814
SQ SEQUENCE 931 AA; 107200 MW; 81AE70094093C023 CRC64;

```

Alignment Scores:

Pred. No.:	6.68e-14	Length:	931
Score:	417.00	Matches:	226
Percent Similarity:	35.78%	Conservative:	125
Best Local Similarity:	23.04%	Mismatches:	350
Query Match:	8.70%	Indels:	282
DB:	1	Gaps:	42

```

US-09-976-674-4 (1-2617) x DAP1_YEAST (1-931)
QY 33 CGCGCGACGAGCGAGCA-GCCGCACAGATACCCGCCCGCTTCAGGTGACG 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 ArgProThrGlnAlaThrIleAspValThrAspValProGlnThrProPheLeuGlnGlu 74
QY 92 AAGCAGCTGTGGAGCGGCTCCGAGCATATCCAGCGAGCGGCGCAAGTACTCGGGCTC 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 GlnTyrSerMetArgProArgArg-----
QY 152 ATTGTCAACAGGCGCCCGACGATTCAGTTGTGCAAGAGAGGATGACTGCGGCC 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 -----GluSerPheGlnPhe-----AsnAspIleGluAsnGlnHis 94

```


QY 212 CACTCCACCGCCTACTACTACCTGGAGATGCATATGACCGACGAGAACTCCCTCTC 271
 |||:||||| :||| :||| :|||:||||| :||| :||| :|||
 Db 95 HistHrHisSerPhePheSerVal----- 102
 QY 272 TACTCTAGATTCACCAAGAGTCCGGAAGAGCGCTGCTCTCTGCTCCGGAAGACG 331
 :||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 103 ---Asn-LysPheAsnArgArgTrpGlyGlu-----TrpSerLeuProGluLysArg 118
 332 ATGCTGCATC----- 341
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 118 gserTyrValLeuValPheThrLeuLeuAlaLeuSerValLeuValLeuValLeu 138
 QY 342 -----ATTCAGGCGACCGCCACCATGAGG-----TC 370
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 138 uilleProSerLysLeuLeuProThrLysLleThrArgProLysThrSerAlaGlyAspSe 158
 QY 371 TACTCTGGGAGAGAGAGCTG-CTGAGGAGCGGAAAGCGCTGGGGGCTCTTCCGATCAC 429
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 158 rserLeuGlyLysArgSerPheSerIleGluAsnValLeuAsnGlyAspPheAlaIlePr 178
 QY 430 CTCTAGACTTCAC-----ACGAGAG 453
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 178 ogLiasPrThPheHisPheIleAsPrProGlnArgLeuLeuGlyLiasSerAspPr 198
 QY 454 TGGCCTCTTCCTCTTC-----CAGGC 474
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 198 ogLysLeuTyrPheThrLysGluIleAspGlyHisThrAsnPheIleAlaLysGluLe 218
 QY 475 CAGCAACAGCCTCTTCACCTGTCGACGCGGCGGCAAGACGCTTCATGATGTCCTCAT 534
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 218 upPheAspGluThrPheGluValAsnLeuGlyGly---AsnArgPheLeuTyrGluGlyVa 237
 QY 535 GAAACCCCTGGAATTCAGACCCAGCTGCTGAGGCCCCGATGAGCCCAAAATCTGCC 594
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 237 Iglu---PheThrValSerThrValGlnIleAsnTyrLysLeuAspLysLeuIlePheG1 256
 QY 595 TGGCCGAC---CCTGCTCTTCTCTCTTCATCATCAATACAGCAGCAGCTTCGTCACAT 651
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 256 yThrAsnLeuGluSerGluPheArgHisSerSerLysGlyPheTyrTrpIleLysAspLe 276
 QY 652 CGAGACAGCGGAG-----GAGCGG----- 670
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 276 uasnThrGlyAsnIleGluProIleLeuProGluLysSerAspAsnTyrGluLe 296
 QY 671 -----CGCTGACCTTCTGCCACCAAGT----- 694
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 296 uGlyLeuSerLysLeuSerTyrAlaHisPheSerProAlaTyrAsnTyrIleTyrPheVa 316
 QY 695 -----TTATCCATGCTCGATGACCCCAAGTGGCGGTGGCCACCTTCGTCATCA 750
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 316 lTyrGluAsnAsnLeuPheLeuGlnGlnValAsnSerGlyValAlaLysLysValThrG1 336
 QY 751 GGAAG---GAGTTCGACCGCTTCAC----- 772
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 336 uaspgLysSerLysAspLlePheAsnAlaLysProAspTrpIleTyrGluGluValLe 356
 QY 773 -----GGTACTGCTGGTGGTGGCCCGACAGCTCTCGGAGAGTTAGAGGCGCT 819
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 356 uAlaSerAspGlnAlaIleTrpTrpAlaProAspAspSer----- 369
 QY 820 CAAGACCTGGGAATCTCGATGAGGAAGTGCAGCGGAGCGGAGGTCATTCACGT 879
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 370 -----LysAlaValPheAlaArgPheAsnAspThrSerValAspAspIleArgLe 386
 QY 880 CCCCTCTCTGCGCTAGAGAAAGAGACGAGC----- 913
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 386 u-----AsnArgTyrThrAsnMetAsnGluAlaTyrLeuSerAspTh 400
 QY 914 -TCGTATCGGTACCCAGAGACGAGCAGCAAGATCCCAAGATTGCTGAAACTGGCTGA 972
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 400 lTysIleLysTyrProLysProGlyPheGlnAsnProGlnPheAspLeuPheLeuValAs 420

QY 973 GTTCAGACTGACAGCCAGGCGCAAGATGCTCTGACCCAGAGAGAGAGCTGTGACGCC 1032
 ||| :||| :||| :|||:||||| :||| :||| :|||
 Db 420 nLeuGlnAsnGlyLleIleTyrSerIleAsnThrGlyGlyLysAspSerIle----- 438
 QY 1033 CTTCAGCTCGCTGTTCGGAAGTGGATACATCGCCAGGCGCGAGTGGACCCGGATGG 1092
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 439 -----LeuTyrAsnGlyLysTrpIleSerProAs 448
 QY 1093 CAATACGCTGGGCGCATGCTCTGAGCCGCGCCAGCAGTGCTCCAGCTGCTCTCT 1152
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 448 pThrPheArgPheGluIle---ThrAspArgAsnSerLysIleLeuAspValLysValTly 467
 QY 1153 CCCCCGCGCCTGTTCATCCCGACAGCAGCAAGAGAGAGAGCGGAGCTCTGCGCAG 1212
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 467 rAsp-----lleProSerSerGln-----MetLeuThrValAr 478
 QY 1213 AGCTGTCCCGAGAGATGTCACCGGTATGTTGTATGAGAGAGGTCAACATCTGTGAT 1272
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 478 gAsnThrAsnSerAsn-----LeuPheAsnGlyTrpI1 489
 QY 1273 C---AATGTTATGACATCTTCTAT-----CCCTCCCAATCAGAGGAGAGAGAGA 1323
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 489 eGluLysThrLysAspIleLeuSerIleProLysProGluLeuLysArgMetAsp-- 508
 QY 1324 GCTCTGCTTCTCCGCGCCATGATGCAAGACCGGCTTCGCCATTTGTCAAAAGTCA 1383
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 509 -TyrGlyTyrIleAspIleHisAlaAspSerArgGlyPheSerHisLeuPheTyrTyrPr 528
 QY 1384 CGCGCTTTTAAATCCAGAGGCTACGATGAGTGAAGTCAAGCCCGGAGAAATGA 1443
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 528 oThrValPhe----- 531
 QY 1444 ATTTAAGTGGCCCATTAAGAGAGATGCTCTGACACGCGTGAAAGGAGTTTGGC 1503
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 532 -----AlaLysGluProIleGlnLeuThrLysGlyAsnTrpGluValThrG1 547
 QY 1504 GAGCAGCGCTCCAAAGATCGTGTCATGAGAGACCAAGTGCGTACTTCACAGGGAC 1563
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 547 y-----AsnGlyLleValGlyTyrGluTyrGluThrAspThrIlePhePheThrAlaAs 565
 QY 1564 CAAGACAGCGCGCTGAGACACACCTCTACGTGTGCTATGAGGCGCGCGAGAT 1623
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 565 nGluIleGlyValMetSerGlnHisLeuTyrSerIleSerLeuThrAspSer----- 582
 QY 1624 CGTACGCTCACACAGCGCGCTTCTCCATAGCTGCTCCATGAGCCAGAACTTGACAT 1683
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 583 -----ThrThrGlnAsnThrPheGlnSerLeuGlnAsnProSerAspLysTyrAspPh 600
 QY 1684 GTTCGTAGCCACTACAGCAGCGTGAAGC----- 1711
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 600 eTyrAspPheGluLeuSerSerSerAlaArgTyrAlaIleSerLysLysGlyProAs 620
 QY 1712 -ACGCG-----CCCTGCGTCAAGCTTCAAGCTGAGCGCGCCGAGCA 1755
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 620 pThrProIleLysValAlaGlyProLeuThrArgValLeuAsnValAlaGluIleHisAs 640
 QY 1756 CGAC-----CCCTGCAACAAGACCGCGCTTCTGGCTGACATGATGAGGAGAC 1806
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 640 pAspSerIleLeuGlnLeuThrLysAspGluLysPheLysGluLysIle----- 656
 QY 1807 CAGCTGCGCGCGATATGTTCTCCGAGAGATCTTCATTTCCACAGCGCC----- 1858
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 657 -----LysAsnTyrAspLeuPro---llePheSerTyrIleThrMetValLeuAs 672
 QY 1859 -TCGATGTGCGGCTCTACGCGCATGATCTACAGCCCGCCAGCTTGCAGCGGAGAGA 1917
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 672 pAspGlyValGluLysValIleGluIleLysProLAsnLeuAsnProLysLysLys 692
 QY 1918 GCACCCCAAGCCTCTTGTATATGAGGCGCCCGAGGTGACGCTGTGAATACTCTT 1977
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 692 sTyrProIleLeuLeuAsnIleTyrGlyGlyProGlySerGln----- 706
 QY 1978 CAAGGCATCAAGTACTTGGCGCTCAACACTGCGCTCCCTGGGCTACGCGGTGTGT 2037

```

Db 707 -----ThrPheThrThyLysSerSerLeuAlaPheGluGlnAlaVala 720
QY 2038 GATTGACGCG-----AGGGGCTCCGTGACGAGG 2067
Db 720 ValSerGlyLeuAspValAlaValLeuGlnIleGluProArgGlyThrGlyGlyVal 740
QY 2068 GCTTCGGTTTCGAGGGGCGCTGAAAAACAATAGGGCAGGTGAGATTCAGGACCAAGT 2127
Db 740 YTPSerPheArgSerThrAlaArgGlyLysLeuGlyTyrThrGluProArgSpleth 760
QY 2128 GGAG---GGCCTCAGTTCGTGGCCGAGAAAGTATGGCTTATGACCTGAGCCGAGTTC 2184
Db 760 rGluValThrLysLysPheIleGlnArgAsnSerGlnHisIleAspGluSerLysIleAl 780
QY 2185 CATTGCATGGTGGTCCAGGAGGGGCTCCCTGGCTCATGAGGGGCTA---ATCCACAAAGCC 2241
Db 780 AlletPrgLysThrSerGlyGlyGlyPheThrSerLeuLysThrValGluLeuAspAsnG1 800
QY 2242 CCAAGTGTTCAGAGTGCATCCGCGGCTGCCCGCTCAGCTGTGATGGCTTACGACAC 2301
Db 800 YAspThrPheLysThrAlaMetAlaValAlaProValThrAsnThrPheLysThrAspSe 820
QY 2302 AGGCTACACTGACGCGCTACACTGACGCTCCCTGAGAACACACAGCAGCGCTATGAGCGCG 2361
Db 820 rValTyrThrGluArgTyrMetAsnGlnProSerGluAsnHisGluGlyTyrPheGluVa 840
QY 2362 TTCGCGTGGCCCTGACGCGTGAAGAAGTGCACCAATGACCCCAACGCGCTGTATTCCTCA 2421
Db 840 lSerThrIleGlnAsnThyLysSerPhe---GluSerLeuLysArgLeuPheIleValHI 859
QY 2422 CGGCTTCCTGAGCAAGAAACGTGCATTTTTCACACAACCTCCCTGCTCCCAACTGAT 2481
Db 859 sGlyThrPheAspAspAsnValHisIleGlnAsnThrPheArgLeuValAspGlnLeuAs 879
QY 2482 CCGAGCAGG---AAACCTTACCACTCCAGATCTACCCCAAGCAGAGACAGACTATTCG 2538
Db 879 nLeuGlyLeuThrAsnThrAsnThrAsnMetHisIlePheProAspSerAspHisSerIleAr 899
QY 2539 C 2539
Db 899 g 899

```

```

RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN-SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STR2 and SGA genes
from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38061; CAA86176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001458; MDC1.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
FT CARBOHYD 874 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2BBD61AA9D CRC64;

Alignment Scores:
Pred. No.: 5,38e-11 Length: 1367
Score: 359.50 Matches: 216
Percent Similarity: 36.08% Conservative: 104
Best Local Similarity: 24.35% Mismatches: 380
Query Match: 7.50% Indels: 187
DB: 1 Gaps: 30

US-09-976-674-4 (1-2617) x AMTH_YEAST (1-1367)
QY 9 CCATGGCCACACCGGAGCCCAACGCGCGAGCGAGCGCCGACAGATGACC 68
Db 272 ProThrProThrThrThrThrThrSerCysThrLysGluLysProThrProThrHisAsp 291
QY 69 CGGCGCGCGCGCTTCACAGTGCAGACACCTCGTGAGCGGCTCCGAGATATCCAGC 128
Db 292 ThrThrProCysThrLysLysThrThrThrSerLysThrCysThrLysLysThrThr 311
QY 129 GCA-----GGCGCAAGTACCGGGGCCATGTCACAAAGCGCCCGCCAGC 173
Db 312 ThrProValProThrProSerSerThrThrGluSerSerAlaProValProThr 331
QY 174 ACTTCAGATTGTGCAAGAAAGAGATGAGTGTGGCCCACTCCACAGCGCTTACTAC 233
Db 332 ProSerSerSerThrThrGluSerSerSer-----AlaProValThrSerSerThr 348
QY 234 TGGGATGCAATATGCAAGCCGAGACACTCCCTCTACTCTG-----AGA 291
Db 349 ThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSer 368
QY 282 TTCACAAAGAGTCCGGAAGAGAGGCTGTGCTCCGCTCT-----GGAAGCAGATGC 335
Db 369 AlaProValThrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThr 388

```

OY	336	TGATCATTTCCAGGCCACGCCCCAC	---ATGGGCTTACTCTGGGAGGAGCTGC	392
Db	389	gluserSerSerAlaProValProThrProSerSerSerThrThrgluserSerSerAla		408
OY	393	TGAGGAGCCGGAACGGCTGGGGGCTCTGGGATACACCTCCATACAGCTTCCACACCGAGA		452
Db	409	ProValThrSerSerThrThrgluserSerSerAlaProValThrgluserSer		426
OY	453	GTGGCCTTCCTCTCTCCAGGCGAGCAACGCCCTTCACATGTGGCGACGGCGAGAGA		512
Db	427	---ThrgluserSerSerAlaProValThrgluserSerThrgluserSerSerAlaPro		445
OY	513	ACGGCTTCATGTGTCTCCCTATGTAACCCGCTGGAATCAAGACCCAGTCTCAGGGCCC		572
Db	446	ValThrSer	---SerThrgluserSerSerAlaPro	457
OY	573	GGATGGACCCCAAAATCGCCCTGGCCGACCGCTCTCTTCCTC	-----TCATCATTA	626
Db	458	ValProThrProSerSerSerThrgluserSerSerAlaProValThrgluserSerThr		477
OY	627	ACAGGACCTGTGGTGGCCCAACATGCAGACAGACAGGAGGCGGGCGAGCTCTCC		686
Db	478	ThrgluserSerSerAlaProValProThrProSerSerSerThrgluserSerSer		497
OY	687	ACCAAGCTTTATCCATATGCTCTGATGACCCCAAGTCTGGGGGTGGCCACTTCGTGA		746
Db	498	AlaProValThrSerSerThrgluserSerSerAlaProValProThrProSer		516
OY	747	TACAGGAAGATTGACCCGCTTCACTGGGTGGTGGTGGCCCAAGCTCTCTGGGAG		806
Db	517	-----SerSerThrgluserSerSerAlaProAlaProThrProSerSer		533
OY	807	GTTCAGAGGCGCTCAAGACCGCTCGATCTCTATGAGAGATCGATAGTCCGAGTGG		866
Db	534	ThrThrgluserSer		538
OY	867	AGGTCAATTCACGTCCTCTCTCGCTGTAAGAAAGAGACGACTCGTACGCTGAC		926
Db	539	-----SerAlaProVal		547
OY	927	CCAGGACGAGCAGCAAGAAATCCCAAGATTGCTTGAACGTGCTGATTCACAGACTGACA		986
Db	548	gluserSerSerAlaProValPro		555
OY	987	GCCAGGGCAAGATGCTGTCGACCCAGAGAGAGAGCTGTGACGCCCTTCAGCTGCTGT		1046
Db	556	-----ThrProSerSerThr		561
OY	1047	TCCCGAAGTGGAGTACATGACGCGAGGCGGGGTGACCCGGGATGGCAATACGCTGGG		1106
Db	562	ThrgluserSerSerThrProValThrgluserSerThrgluserSerSerAlaProVal		581
OY	1107	CCATGTTCTCTGGACCGGCCCCACAGTGGCTCAGCTGCTCTCTCC	-----CCC	1157
Db	582	ProThr	-----ProSerSerSerThrgluserSerSerAlaProValPro	597
OY	1158	CGGCGCTGTTCATCCGAGACAGACAGAAATGAGAGGAGGCGGTACCTCTGCGACAGCTG		1217
Db	598	ThrProSerSerSerThrgluser	-----SerSerSerAlaProAlaPro	612
OY	1218	TCCCCAGGAATGTCACGCCCTATGTGTGAGTACAGAGAGGTCAACCAAGCTGTGATCATG		1277
Db	613	ThrProSerSerSerThrgluser	-----SerSerSerAlaProValThrgluserSerThr	630
OY	1278	TTTCATGACATCTCTATCCTTCCCCCAATCAGAGGAGGAGAGAGCTCTGCTTCTCC		1337
Db	631	ThrgluserSerSerAlaProValProThrProSerSerSerThrgluserSerSer		650
OY	1338	GGCGGCAATGATGCAAGACGGGCTCTGGCATTTTGACAAATCAACCGCGCTTTAAAT		1397
Db	651	AlaProValProThrProSerSerSerThrgluserSerSerAlaProValProThr		670

QY	1398	CCGAGGCTACGANTTGAGAGCCGCTTCAGCCCGCGGGAACATGATTTAAAGTCCCA	1457
Dd	671	ProSerSerSerThrThrGluSerSerAla-----	681
QY	1458	TTAAGAAAGATTGCTCTGACCAAGCGGTGAATGGAGAGTTTGGCAGCAGCCTCA	1517
Dd	682	-----ProValThrSerSerThrThrGluSerSerAlaProValProThrProSerSer-----	712
QY	1518	AGATCTGGGTCATGATGAGAGACCAAGCTGGTACTTCCAGGGCACCAAGACGCGC	1577
Dd	695	ValThrSerSerThrThrGluSerSerAlaProValProThrProSerSer-----	722
QY	1578	TGGGAGCAACCTCTACGTGTAGCTATGAGGGCGCGGAGATCTAGCCTCACA	1637
Dd	713	---SerThrThrGluSerSerAlaProValProThrProSerSerThrThrGlu	731
QY	1638	CGCCCGGCTCTCCCATAGCTGCTCCATGAGCCAGACACTTCACATGTTCCGACGCCACT	1697
Dd	732	SerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerAlaPro	751
QY	1698	ACAGCAGGTGAGCAGCGCCCGTGGTCACGCTTCAACGCTGACGCGCCCGCAGCAG	1757
Dd	752	ValThr-----SerSerThrThrGluSerSerAlaProValPro	765
QY	1758	ACCCCTGACACAGCAGCCCGCTTCTGGGCTAAGCATGTAGAGCAGCAGCTCCGCC	1817
Dd	766	ThrProSerSerSerThrThrGluSerSerAlaProValProThrProSerSer---	784
QY	1818	CGGATTATTGTTCCCTCCAGAGATCTTCCATTCCACACGCGCTGGAGTGGGGCTACG	1877
Dd	785	-----SerThrThr	787
QY	1878	GCATGATCTACAAAGCCCGCTTCACAGCAGGAGACCCACCGCTCCCTTTG	1937
Dd	788	GluSerSerSerAlaProValProThrProSerSerThrThrGluSerSerAla	807
QY	1938	TATTTGAGGCCCCCAGGTGACCTGCTGAATACTCTTCAAGGATCAAGTACTTC	1997
Dd	808	ProValProThrProSerSerSerSerAlaThrSerSerAlaProSerSerThrPro	827
QY	1998	GGCTCAACACACTGGCT-----CCCTGGGCTACCGCGTGGTGTGATTGACG	2045
Dd	828	PheSerSerSerThrGluSerSerSerAlaProValProThrProSerSerSerThrThr	847
QY	2046	-----CCAGGGGCTCTCTTCACGCGAGGCTTC-----GGTTCGAAGGCGCCCTGAAA	2093
Dd	848	GluSerSerSerAlaProValSerSerSerThrThrGluSerSerValAlaProValPro	867
QY	2094	ACCAAAATGGCCAGGTGAGATCGAGACCAAGGTGAGGGCTCGCATGTTGTGGCCGAGA	2153
Dd	868	Thr-----ProSer	870
QY	2154	AGTATGGCTTCATCGACCTACGCGGAGTTGCCATCATGAGCTGGCTACAGGGGGCTTC	2213
Dd	871	SerSerSerSerAlaThrSerSerSerAlaProSerSerIleProPheSerSerThrGlu	890
QY	2214	TCTGCTCATGGGCTATCATCAACAGCCAGATGTTAGGTGGCA-----TCCGGAGTG	2270
Dd	891	SerPheSerThrGlyThrThrValThrProSerSerSerLysTyrProGlySerGlnThr	910
QY	2271	CCCCGGTACCGCTGTGATGGCTACGACAGAGGATACAGT-----AGCGCTACATGG	2334
Dd	911	GluThrSerSerSer-----ThrThrGluThrIleValProThrProLysThrThr	928
QY	2335	ACGTCCCTGAGAACACACAGCAGGCTATGAGCGGGTTCGTGGCCCTGCACGTGGAGA	2384
Dd	929	ThrSerValThrThrProSerThrThr-----	937
QY	2385	AGCTGCCAATGAGCCCAACCGGCTCTATATCTCCACAGGGCTTC-----TGGACG	2435
Dd	938	-----ThrIleThrThrThrValLysSerThrGlyThrAsnSerAlaGlyGluThrThr	955
QY	2436	AAAAGCTGACCTTTTCCACACAACTTCGTGCTCCCAACTGATTCGAGCAGGAGAAC	2495


```

FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1351 1351 H -> L (IN REF. 3).
FT CONFLICT 1412 1412 T -> S (IN REF. 3).
FT CONFLICT 1449 1449 L -> P (IN REF. 3).
FT CONFLICT 1504 1504 M -> T (IN REF. 3).
FT CONFLICT 4192 4192 G -> S (IN REF. 2).
SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;

```

```

Alignment Scores:
Pred. No.: 6.84e-07 Length: 5179
Score: 279.50 Matches: 210
Percent Similarity: 31.85% Conservative: 113
Best Local Similarity: 20.71% Mismatches: 345
Query Match: 5.83% Indels: 347
DB: Gaps: 43

```

US-09-976-674-4 (1-2617) x MUC2_HUMAN (1-5179)

```

QY 18 CCACGGGACCCCAAGCGCCGAGCGAGC-----CAG 53
DB 1197 ProProGluAlaSerValProThrCylsLysSerCysValCysThrAsnSer 1216
QY 54 CCGCACAGATGACCGCGCCGCTCCAGGTCAGAGACACTCGGAGGCGCTCC 113
DB 1217 SerGlnValValCysArgProGluGlnLysIleLeuAsnGlnThrGlnAspLysAla 1236
QY 114 -----GGAGCATCATCCAGCGACGCGGCAAGTACTCGGGCTCA 152
DB 1237 PheCysTyrTrpGluIleCysGlyProAsnGlnThrValGlnLysHisPheAsnIleCys 1256
QY 153 TTGTCAAGAAGCGCCCAACGACTTCCAGTTGTGCAGAAGACGATAGCTGGCCCC 212
DB 1257 SerIleThrThrArgProSerThrLeuThrPhe---ThrThrIleThrLeuProThr 1275
QY 213 ACTCCACCGCTCTACTACTGGAATGCATATGCGAGCGAGAGAACTCCCTCTCT 272
DB 1276 ThrProThrSerPheThrThr-----ThrThrThrThrThrThrProThrSer 1291
QY 273 ACTGTGATTTCCCAAGAAGTCCGGAAGAGCTCTGCTGCTCTCTGGAAGCAGA 332
DB 1292 SerThrValLeu-----SerThrThrProLysLeuCysCys----- 1303
QY 333 TGTCTGATC-----ATTTCAGCGCCAGCCCAACCATGGGCTCTACTCTGGGAGAG 386
DB 1304 LeuThrPserAspTrpIleAsnGlnAspHisProSerSerGlySerAspAspLysArg 1323
QY 386 ----- 386
DB 1334 GluProPheAspGlyValCysGlyAlaProGluAspIleGluCysArgSerValLysAsp 1343
QY 387 -----ACCTGCTGAGGAGC----- 401
DB 1344 ProHisLeuSerLeuGlnHisIleGlnLysValGlnCysAspValSerValGlyPhe 1363
QY 402 -----GGAAC----- 407
DB 1364 IleCysLysAsnGlnAspGlnPheGlnLysAsnGlnLysProPheGlyLeuCysTyrAspTyrLys 1383
QY 408 -----GCCTGGGGG-----TCTTCGGCATCACTCTCTAGC 437

```

```

DB 1384 IleArgValAsnCysCysTrpProMetAspLysCysIleThrThrProSerProProThr 1403
QY 438 ACTTCACAGCGAGAGTGGCTTTCCTTCACAGGCCAGCAACAGCTCTCCACGTGC 497
DB 1404 ThrThr-----ProSerProProProThrThr 1413
QY 498 GCGAGCGGGGAGAGCGGTCATGTCGTCGCCATATGAACCGGTGAATCAAGACC 557
DB 1414 ThrThrLeuProProThrThrThrProSerPro----- 1424
QY 558 AGTGCTACGGGCGCCGATGAGCCCAAAATCTGCCGCGACCCCTCTCTCTCTCT 617
DB 1425 -----ProThrThrThrThrThrThrProProProThrThrThrProSerPro 1440
QY 618 TCATCATACAGCAGCAGCTGTGGGTGCGCAACATCGACAGCGGAGCGCGGCTGA 677
DB 1441 ProIleThrThrThrThrThrProLeuProThrThrThrProSerPro----- 1456
QY 678 CCTCTCGCCACCAAGTTTATCCAAATGTCGTGATGACCCCAAGTCTCGGGGTGCGCA 737
DB 1457 ProIleSerThrThrThrThrProProProProThrThrThrProSer-----Pro 1472
QY 738 CCTTCGTATACAGAGAGTTCGACCGCTTCACCTGAGTACTGAGTGGCCCAAGCCT 797
DB 1473 ProThrThrThrProSerProProThrThr-----ProSerPro 1486
QY 798 CCTGGGAAGTTACAGAGGCGCTCAAGACGCTGCAATCTGTATGAGAGAGTCAGT 857
DB 1487 Pro-----ThrThrThrThrThr 1493
QY 858 CCGAGTGTGAGTCAATTACGTCCTCTCTGCGCTTGAAGAAAGAGAGAGTCTGT 917
DB 1494 ProProProThrThrThrProSerProProMet-----ThrThrPro 1507
QY 918 ATCGCTACCCAGCAGCAGCAGCAAGAAATCCCAAGATTG-----CCTTGAAC 965
DB 1508 IleThrProProAlaSerThrThrThrLeuProProThrThrThrProSerProProThr 1527
QY 966 TGGCTGACTCCAGACTGACAGCAGCAGCGCAAGATCTGCAACCCAGAGAGAGCTGG 1025
DB 1528 ThrThrThrThrThrProProProThrThrThrThrProSerProProThrThrProIle 1547
QY 1026 TGCAGCCCTTCAGCTCGGTGTCGCGAGTGGAGTACATGCGCGAGCGC----- 1076
DB 1548 ThrProProThrSerThrThrThrLeuProProThrThrThrProSerProProThr 1567
QY 1077 -----GGTGACCCGCGGATGGCAATACGCTGCGGCGATGTCCTGAGCGCGCCAGC 1130
DB 1568 ThrThrThrThrProProProThrThrThrThrProSerProProThrThrThrProSer 1587
QY 1131 AGTGCTCAGCTGCTCTCTCTCCCGCGCGCTGTTCATCCGAGCAGAGAAATGAG 1190
DB 1588 ProProThrIleThrThrThrThrProProProThrThrThrProSerProProThr 1607
QY 1191 AGCAGCGGCTAGCCTTCGTCGAGAGCTGCCAGGAAGTCCAGCGGATGAGTGTACG 1250
DB 1608 ThrThrThrThrProProProThrThrThrThrProSerProProThrThrThrProIleThr 1627
QY 1251 AGGAGTCAACCAAGCTGTGATCAATGATGATCATGATCTGTATGCTTCCCAATCAG 1310
DB 1628 -----ProProThrSerThrThrThrLeuProProThrThrThrProSerProPro 1645
QY 1311 AGGAGAGAGC-----AGCTCTGCTTCTCCGCGCAATGACAGACCG 1358
DB 1646 ThrThrThrThrThrProProProThrThrThrThrProSerProProThrThrThrPro 1665
QY 1359 GCTTCGCAATTGTACAAAGCA-----CCGCGGTTTAAATCCAGGCGTACAGTT 1412
DB 1666 SerProProIleThrThrThrThrThrProProProThrThrThrProSerSerProIle 1685
QY 1413 GGAGTGAAGCCCTTCAGCCCGGGAAGATGAATTAAGTGCCTTAAGAGAGATTTG 1472

```

```

Db 1686 ThrThrProSerProPro-----1692
Qy 1473 CTCGACACAGCGGTGAATGGAGAGTTTGGCGAGCAGCGCTCCAGATCTGGGTCAATG 1532
Db 1692 -----1692
Qy 1533 AGGAGACCAAGCTGTGTTACTTCCAGGGCACCAGAGACGCGCTGGAGACACCTCT 1592
Db 1693 -----1696
Qy 1593 ACGTGTGACGTATGAGGGCGCGGAGATGATGCGCTGACGACGCGCGCTTCGCC 1652
Db 1697 Thr-----1704
Qy 1653 ATAGCTGTCATGAGCAGACACTTCGACATGTCGACGACCTACAGCAGCGTGAGCA 1712
Db 1705 ProSerSerProIleThrThrThrThrProSerSerThrThrThrProSer-----1722
Qy 1713 CGCGCGCTGCGTCACTGCTACAGCTGAGCGGCGCGGACGACGACCGCC-----1763
Db 1723 -----ProProProThrThrMetThrThrProSerProThrThrThrProSerProPro 1740
Qy 1764 -----1793
Db 1741 ThrThrThrMetThrThrLeuProProThrThrThrThrSerSerProLeuThr-----1757
Qy 1794 TGATGAGAGCAGCAGCTGCGCGCGGATTTGTTCCAGAGATCTTCATTTCCACA 1853
Db 1758 -----1775
Qy 1854 CGCGCTGCGATGCGCGCTACGCGATGATCTACAGCGCGCGCTTCGAGCGCAGGA 1913
Db 1776 -----1782
Qy 1914 AGAGACACCCAGCTGCTCTTTTATATGAGGCGCGCGAGCTGCTGTAATACT 1973
Db 1783 -----1789
Qy 1974 -----1997
Db 1790 GlyThrLeuAspSerGlyGlyProAsnPhenIleHisProGlyGlyAspThrGluLeuIle 1809
Qy 1998 GCGTCAACACACCTGCGCTCCCTGCGCTACGCGCGTGTGATGAGCGCGCGGCTCT 2057
Db 1810 GlyAspValCysGlyProGlyTTPAlaAla-Asn-----1821
Qy 2058 GTCGACGAGGCTGCGTTCGACAGGGCGCCGAAACCAATGGCGCAGTGA-----2112
Db 1821 eSerCysArgAlaThrMetTyr-----ProAspValProIleGlyGlnLeuGlyGlnTh 1839
Qy 2113 -----2129
Db 1839 rValValCysAspValSerValGlyLeuIleCysLysAsnGluAspGlnLysProGlyG 1859
Qy 2130 AGC-----2140
Db 1859 yValIleProMetAlaPheCysLeuAsnTyrGluIleAsnValGlnCysCysGlnCysVa 1879
Qy 2141 TTGCTGGCCGAGAGTATGCTTCATGACCTGAGCCGAGTGGCATTCAGGCTGCTCC 2200
Db 1879 lThrGlnProThrThrThrThrThrThrThr-----ThcGlnAsnProThrProProThrTh 1898
Qy 2201 TACGGGGCTCTCTGCTGCTCATGGGCTAATCCACA-----AGCCCAAGGTTCAGAGTG 2257
Db 1898 rThrProIleThrThrThrThrThrValThrProThrProThrProThrGlyThrGlnTh 1918
Qy 2258 GCGATCGCGGGTGC-----CGTCAACGCTGAGTGCCTGACGACAGAGG 2305
Db 1918 rProThrThrThrProIleThrThrThrThrThrValThrProThrProThrProThrGln 1938
Qy 2306 TACACTAGCGCTCATGAGAGTCCCGAGAGAACAGACGACGCGCTATGAGCGGGTTCC 2355
Db 1938 yThrGlnThrProThrThrThrProIle-----ThrThrThrThrThrValThrProThrPr 1957

```

```

Qy 2366 GTGGCCCTGC-----ACGTGAGAGCTGCCAATGAGCCCAACCGCTTCATC 2416
Db 1957 oThrProThrGlyThrGlnThrProThrThrProIleThrThrThrThrThrValTh 1977
Qy 2417 CTCACGCGCTTCTGAGCAGAAACGTGCACT-----TTTCCACACAAAC 2461
Db 1977 rProThrProThrProThrGlyThrGlnThrProThrThrProIleThrThrThrTh 1997
Qy 2462 TTCCTGCTGCCAATGATGATCCAGAGGGAACCTTACGAGCTCGATCTACCCGAC 2521
Db 1997 rThrValThrProThrProThrProThrGlyThrGlnThrProThr-----ThrThrProIl 2016
Qy 2522 GAGAGACACAGTATTCGCTGCGCGCGAGCTGCGGCGAGCACT 2561
Db 2016 eThrThrThrThrThrThrValThrProThrProThrProThrProThr 2029
RESULT 15
YH1_EBV STANDARD; PRT; 660 AA.
ID YH1_EBV
AC P03181;
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
OS Hypothetical BHLF1 protein.
OC Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, not RNA stage; Herpesviridae;
OX Gammaherpesvirinae; Lymphocryptovirus.
RN NCBI_TaxID=10377;
RX MEDLINE=64270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Telford P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01555; NOT_ANNOTATED_CDS.
DR PIR: A03742; Q0BE3.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;
Alignment Scores:
Pred. No.: 2,38e-06 length: 660
Score: 265.50 Matches: 218
Percent Similarity: 32.16% Conservative: 57
Best Local Similarity: 25.50% Mismatches: 283
Query Match: 5.54% Indels: 297
DB: 1 Gaps: 49
US-09-976-674-4 (1-2617) x YH1_EBV (1-660)
Qy 142 CTCGCGCTCATTTGTCAACAAGC-----GCCCCACAG-----CTTCAAGTTTGTGCA 189
Db 1 MetGlyThrProCysGlnSerAlaArgGlyProArgThrThrProLeuProHisCys---19
Qy 190 GAAGACGAGTGAAGTGTGGGCGCCGACCTCCAGCGCTTACTAGTGGGATGGCATTTGG 249
Db 20 -----ProProProCysLeuProGlyAlaProAspGln 30

```

[illegible]

QY	1246	GTACGAGGAGGTCCACAACGCTTGATCAATTGCATGACATCTTCTATTCCTTCCCACA	1305
Db	335	-----ProglYgLYAlaValProSerGlyAlaThrProHisPro-	339
QY	1306	ATCAGAGGAGGAGGAGCGTCGCTTTTCGCCGCCAATGAATGGAA	1355
Db	340	-----GluArgGLySeriLProAlaAspProAlaAlaIleArgLeuProProG	357
QY	1356	-----CCGGCTTCGCACTTTGTTCACAAAGTCAC	1383
Db	357	LurArgInGLuPProArgLeuProGlnAspLeuAlaIleAlaGlnArgCysProAlaGLYP	377
QY	1384	CGCCGTTTTAAATCCAGGGGTAACGATTGGAGTAGACCCTTCAGCCCCGGGAAGATGA	1443
Db	377	roPro-----ProThrArgSeriGLy-----ProThrArgSeriGLy-----	383
QY	1444	ATTTAAGTCCCATTTAAGAAGATGTGCTGACCACGCGGATGGAGATTGGGAGTTTGGC	1503
Db	383	-----	383
QY	1504	GAGGACGCGCTCCAAGATCTGGGTCATGTAGAGAGACCAAGCTGTGTACTTCCAGGCA-	1566
Db	384	-----AlaIleAlaGlnArgThrIleAspArgProProGLYCysProAlaArgSerAla	401
QY	1563	-----CCAGGACAGCGCGCTGGAGCACACCTCTACGTCGTCAGCTATGA	1608
Db	401	rGaSPProGLYCysProArgThr-----ThrPArgTrArgSer-----GlyAlaGln	417
QY	1609	GCGCGCGCGGAGATCGTAGCCTCACACACGCCCGCTTCCCATAGCTGCTCCATGAG	1668
Db	417	rGLYLHISProProProGLYAlaGLYlneArgProSerGLYProThrGLYGLY-----	434
QY	1669	CCAGACTTCGCATGTTTCGTCAGCCACTACAGCACGTCGAGAGAGCGCCCGCTGGTGCA	1728
Db	435	-----ArgPro-----	436
QY	1729	CGTCTACAGCTGAGCGGCCCCGACGACGACCCCTGCACAGACAGCCCGCTTGCGGC	1788
Db	437	-----AlaIleProGLYAla-ProGLYTyrProAlaIleArgProGLYProGLY	451
QY	1789	TAGCATGTATGAGGACGACGACTGCCCCCGGATTATGTTCTCCAGAGATTTTCCATT	1848
Db	452	-----GlyGLYAlaIleValProSerGLYAla-----ThrProHis	463
QY	1849	CCACAC-----GCGCTCGATGTGGCGCTCTACGGCATGATCTA	1887
Db	464	ProGluArgGLYSeriGLYProAlaAspProProAlaIleAlaIleArgLeuProProGLUArg	483
QY	1888	CAAGCCCCACGC-----CTTGACCCAGGGAAGAAGACACCCACGCTGCTTT	1935
Db	484	GLnGLuProArgLeuProGlnAspLeuAlaIleAlaIleArgCysProAlaGLYPro---	502
QY	1936	TGTATATGAGAGGCCCCCAGGTGCAGCTGGTGAATACTCTTCAAAGGATCAAGTACTT	1995
Db	503	-----ProProThrArgSeriGLY-----Ala	509
QY	1996	GCGGCTCAACACT-----GGCTCGCGGGGCTACGCGCT-----GCT	2034
Db	510	AlaIleAlaGlnArgThrIleAspArgProProGLYCysProArgSerAlaArgAsnProGLY	529
QY	2035	TGTGATGTAGCGGCGGCGCTCTGTCCAGCGAGGCGCTCGGTTGCGAAGGCG-----	2085
Db	530	Cys-----ProArgThrTrpArgArgArgSeriGLYAlaGlnArgGLYHisProPro	546
QY	2086	-----CCTGAAAAACCAATGGCCACGCTGAGATCGA-----	2118
Db	547	ProGLYAlaGLYlneArgProSerGLYProThrGLYAlaArgProAlaIleArgProGLYAla	566
QY	2119	-----GGACAGCGTGGAGGGCGCTCGAGTTCGTGCGCGGACAACTA	2157
Db	567	ProGLYTyrProAlaIleArgProGLYGLYGLYAlaIleVal-----	581
QY	2158	TGGCTTATCGACCTTGAGCCGAGTGGCATCCATGGCTGTGCTTACGGGGCTTCCTGC	2217

```

Db      582  -----  |||  ::|||  |||  |
          GCTCATGGGCGTATCCACAAGCCCGCAGGTGTTCAAGGTGGCA-----T  2262
Qy      2218  |||  ::|  |||  |||
          |||  ::|  |||  |||
Db      591  rgllyserglyproialaaspproprolaAlaAlaLarLeuProProgluArgIngluP  611
Qy      2263  CGCGGCGTCCC-----CGGTACCCGTTGTGATGGCCCTACGACAC  2301
          |||  ::|||  |||  |||  |||  |||  |||
Db      611  roArglLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProThrA  631
Qy      2302  AGGGTACACTGAGCGGTACATGAGCGCCCTGAGCAACACAGC-----ACGGCTAAGA  2355
          ::|||  |||  |||  |||
Db      631  rgselGlyAlaAlaAlaGlnArgThrHisArgLrProProglyCysProArgSerAlaA  651
Qy      2356  GCGGCGTTCCTCGGCCCTGCACGTGGAGA  2384
          |||  |||  |||  |||  |||  |||
Db      651  rGasnProGlyCysProArgThrTyrArg  660

```

Search completed: December 12, 2002, 11:49:17
Job time : 72.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 12, 2002, 11:45:57 ; Search time 20.5 seconds
(without alignments)
7512.174 Million cell updates/sec

Title: US-09-976-674-4
Perfect score: 4795
Sequence: 1 caagcttaccatgagccaccca.....ttctgagcgccgcgcatccg 2617

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 segs, 29422922 residues
Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09976674/rnat_04122002.162400.6034/app_query.fasta_1.2759
-DB=Issued_Patents_AA -QEMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=humand40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTMT=plc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09976674.ecgn_1.13_etrnat_04122002.162400.6034 -NCPU=6 -ICPU=3
-NO_XLPEXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1093.5	22.8	310	4	US-09-794-236-4
2	489.5	10.2	759	5	PCT-US93-07923-2
3	489.5	10.2	766	1	US-08-230-491A-3
4	489.5	10.2	766	1	US-08-619-280A-3
5	489.5	10.2	766	2	US-08-940-391-3
6	489.5	10.2	766	4	US-09-794-236-1
7	482.5	10.1	755	5	PCT-US93-07923-3
8	458	9.6	771	4	US-09-462-284-2
9	426	8.9	760	1	US-08-230-491A-2
10	426	8.9	760	1	US-08-619-280A-2
11	426	8.9	760	2	US-08-940-391-2
12	259	5.4	960	4	US-09-219-849-5

c 13	253	5.2	1064	1	US-08-642-255-62	Sequence 62, Appl
14	249.5	5.2	2972	4	US-09-579-181-2	Sequence 2, Appl
15	249	5.2	657	4	US-09-355-166-1	Sequence 1, Appl
16	248.5	5.2	3118	4	US-09-579-181-1	Sequence 1, Appl
c 17	245.5	5.1	1065	1	US-08-642-255-72	Sequence 72, Appl
18	240.5	5.0	632	4	US-09-016-080-1	Sequence 1, Appl
c 19	240.5	5.0	1024	3	US-08-931-820-2	Sequence 2, Appl
c 20	238.5	4.9	1057	3	US-08-931-820-1	Sequence 4, Appl
c 21	238.5	4.9	1366	3	US-08-963-825-19	Sequence 19, Appl
c 22	238.5	4.9	1366	4	US-09-500-811-19	Sequence 19, Appl
c 23	238.5	4.9	1366	4	US-09-570-573-19	Sequence 19, Appl
c 24	238.5	4.9	1366	4	US-09-548-608-19	Sequence 19, Appl
c 25	238	4.9	1442	2	US-08-316-650-12	Sequence 12, Appl
c 26	238	4.9	1442	5	PCT-US95-02251-12	Sequence 12, Appl
c 27	234.5	4.9	1366	4	US-09-585-887-10	Sequence 10, Appl
c 28	234.5	4.9	1366	4	US-09-289-578-10	Sequence 10, Appl
c 29	234	4.9	882	4	US-09-413-814-78	Sequence 78, Appl
c 30	234	4.9	1213	4	US-09-413-814-79	Sequence 79, Appl
c 31	232.5	4.8	762	1	US-08-642-255-120	Sequence 120, App
c 32	232.5	4.8	762	1	US-08-397-633A-31	Sequence 31, Appl
c 33	232	4.8	1185	4	US-09-041-886-23	Sequence 23, Appl
c 34	231	4.8	1461	4	US-09-585-887-9	Sequence 9, Appl
c 35	231	4.8	1461	4	US-09-289-578-9	Sequence 9, Appl
c 36	230.5	4.8	1057	3	US-08-931-820-1	Sequence 1, Appl
c 37	230	4.8	762	1	US-08-642-255-120	Sequence 120, App
c 38	230	4.8	1057	3	US-08-397-633A-31	Sequence 31, Appl
c 39	229	4.8	1057	3	US-08-931-820-1	Sequence 1, Appl
c 40	229.5	4.8	1078	3	US-08-963-825-21	Sequence 21, Appl
c 41	229.5	4.8	1078	4	US-09-500-811-21	Sequence 21, Appl
c 42	229.5	4.8	1078	4	US-09-570-573-21	Sequence 21, Appl
c 43	229.5	4.8	1078	4	US-09-548-608-21	Sequence 21, Appl
c 44	229	4.7	1418	3	US-08-963-825-20	Sequence 20, Appl
c 45	229	4.7	1418	4	US-09-010-999-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-794-236-4
Sequence 4, Application US/0974236
Patent No. 6337069
GENERAL INFORMATION:
APPLICANT: Grouxman, Eric
APPLICANT: Lacroix, Jean-Silvain
APPLICANT: Monod, Michel
TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
FILE REFERENCE: 81985/276823
CURRENT APPLICATION NUMBER: US/09/794,236
CURRENT FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 310
TYPE: PRT
ORGANISM: Homo sapiens
US-09-794-236-4

Alignment Scores:
Pred. No.: 1.08e-76
Score: 1093.50
Percent Similarity: 69.34%
Best Local Similarity: 57.31%
Query Match: 22.81%
DB: 4
Gaps: 1

US-09-976-674-4 (1-2617) x US-09-794-236-4 (1-310)

OY 1553 TTCACGGGACACAGGACACCGCTGAGACACACCTCTACGCGTACGTATAGACGC 1612
||||:|||||
Db 1 PhehtugltyrrlyaspsrPrleuglwhshlsleuryValValserTyValasn 20
|||||:|||||
OY 1613 GCCGCGGAGATCGTACGCTCACCACGCGCGCTTCGCCATAGCTGCATAGACGAG 1672
|||||:|||||

[illegible]

RESULT 2
PCT-US93-07923-2
: Sequence 2, Application PC/TUS9307923
: GENERAL INFORMATION:
: APPLICANT: Morimoto, Chikao
: APPLICANT: Schlossman, Stuart F.
: APPLICANT: Tanaka, Toshiaki

```

1 TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
2
3 NUMBER OF SEQUENCES: 16
4
5 CORRESPONDENCE ADDRESS:
6
7 ADDRESSEE: Fish & Richardson
8
9 STREET: 225 Franklin Street
10
11 CITY: Boston
12
13 STATE: Massachusetts
14
15 COUNTRY: U.S.A.
16
17 ZIP: 02110-2804
18
19 COMPUTER READABLE FORM:
20
21 MEDIUM TYPE: 3.5" Diskette, 1.44 MB
22
23 COMPUTER: IBM PS/2 Model 50Z or 55SX
24
25 OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
26
27 SOFTWARE: Wordperfect (Version 5.0)
28
29 CURRENT APPLICATION DATA:
30
31 APPLICATION NUMBER: PCT/US93/07923
32
33 FILING DATE: 19930819
34
35 CLASSIFICATION:
36
37 PRIOR APPLICATION DATA:
38
39 APPLICATION NUMBER: 07/934,162
40
41 FILING DATE: 21-AUG-1992
42
43 APPLICATION NUMBER: 07/832,211
44
45 FILING DATE: 06-FEB-1992
46
47 ATTORNEY/AGENT INFORMATION:
48
49 NAME: Fraser, Janis K.
50
51 REGISTRATION NUMBER: 34,819
52
53 REFERENCE/DOCKET NUMBER: 00530/055002
54
55 TELECOMMUNICATION INFORMATION:
56
57 TELEPHONE: (617) 542-5070
58
59 TELEFAX: (617) 542-8906
60
61 TELEX: 200154
62
63 SEQUENCE CHARACTERISTICS:
64
65 LENGTH: 759
66
67 TYPE: amino acid
68
69 STRANDEDNESS:
70
71 TOPOLOGY: linear
72
73 PCT-US93-07923-2

```

Alignment Scores:	
pred. No.:	1,04e-29
Score:	489.50
Percent Similarity:	36.27%
Best Local Similarity:	22.43%
Query Match:	10.21%
DB:	5
US-09-976-674-4 (1-2617) x PCT-US93-07923-2 (1-759)	
	length: 759
	Matches: 201
	Conservatives: 124
	Mismatches: 330
	Indels: 241
	Gaps: 35

OY	149	CTCAATTGTCAGAGGGGGCCCGACGACTTCACGTTTGTCAGAAAGGAGTGAATCTGGG	208
Db	19	ValLeuLeuAnaLysGly-----	ThraspAspAlaThr 29
OY	209	CCCCACTCCACGGCCCTACTACTCGGAAATGCCATATGCGACGCGAGAACTCCCTC	268
Db	30	AlaAspSerIarGlySerThrTyThrLeu-----	38
OY	269	CTCTACTGTGAAATCCCAAGAAAGCGTCCGAAAGAGCGCTGCTGCTCTCTCTCGGAAG	328
Db	39	-----ThrAspTyrLeuLysAsnThrTyrIarGleuLysLeuTyrSerLeuAarGTrp---	55
OY	329	CAGATGCTGCATCATTTCCAGGCGCAGCCCGCCACCAGTGGGTCTACTCTCGGAGAGAGAG	388
		:::	:::
Db	56	-----IleSerAspIleGluTyrLeuTyrGlnGluAsn	68
OY	389	CTGCTG-----AGGAGCGGAAACGCTG	412
		:::	
Db	69	IleLeuValPheAsnAlaGluTyrGlyAsnSerSerValPheLeuGluAsnSerThrPhe	88
OY	413	GGGGCTTGGGCATCCACTCTACGATTTCCAC---AGCGAGATGGCGCTTCCTCTTC	469
Db	89	AspAlupPheGlyHisSerIleAsnAspTyrSerIleSerProAspGlyGluPheIleLeu	108

QY 470 CAGGCGACGACAGCCTCTTCACCTGTGCGACGCGGACGACGAGGCTTCATGCTGCC 529
 Db 109 LeuLuluyrAsnYrValyGlnPrpAghIsSerYrThrAlaSerYrAlaSerYr 128
 QY 530 CCAATGAAACCGGTGAAATCAAGACCGCTCAGGGCCCGGATGACCCCAAAATC 589
 Db 129 AspLeuSnYrArgGlnLeuIleThrGlnLulArgIleProAsnAspThrGlnPrpAl 148
 QY 590 TGCCCTCCGACCGCTGCCCTTC---TTGCTCTTCATCATATACAGCAGCCTTGGTGGCC 646
 Db 149 ThrTrpSerProValGlnHisLysLeuAlaValIleProAsnAspIleYrValLys 168
 QY 647 AACATCGACAGACGACGAGGACGCGGCTGACCTTCCTCCACCAAGTTTATCCAAATGTC 706
 Db 169 IleGlnProAsnLeuProSerYrThrArgIleThrTrp-----ThrGlyLysGlnAspIle 186
 QY 707 CTGATGACCCCAAGCTGCGGTGGCCACCTTCGTACATACAGGAAG---TTGAC 763
 Db 187 IleYrAsn-----GlyIleThrAspPrpValYrGlnGlnLulValPheSer 202
 QY 764 CGCTTCACTGGGTACTGGTGGCCCGACACCTTCCTGGGAAGTTCCAGAGGCGCTCAAG 823
 Db 203 AlaYrSerAlaLeuYrTrpSerProAsnGlyThrPhe----- 215
 QY 824 ACCTGCGAATCTGTATGAGGAAGTCGATGATCCGAGTGGAGTCATTCAC----- 877
 Db 216 -----LeuAlaYrAlaGlnPheAsnAspThrGlnLulValIleGlnYrSer 232
 QY 878 ---GTCCCTCTCGCTGCGCTAGAAAGAAAGACGACGCTGTATCGGTACCCGACGACA 934
 Db 233 PheYrSerAspGlnSerLeuGlnYrProYrThrValArgValProYrProYrAla 252
 QY 935 GCGACGACGAATCCCAAGATTGCTTGAACCTGGCTGAGTTCCACAGCTGACAGCCAGGCGC 994
 Db 253 GlyAlaValAsnProThrVal-----LysPhePheValAlaAsnThrAspSerLeuSer 270
 QY 995 AAGATCGCTCGACCCAGACGAGGAGGAGCTGCGACGCTTCGACCTGCTGTCCCGAAG 1054
 Db 271 SerAlaThrAsnAlaThrSerIleGlnIleThrAlaProAlaSerMetLeuIleYrAsp 290
 QY 1055 GTGAGTACATGCCGAGGCGCGGTGACCCGCGGATGGCAATACGCTGGGCGCATGTTC 1114
 Db 291 ---HisYrLeuCysAspValThrTrpAlaThrGlnGlnLulArgIleSer----- 305
 QY 1115 CTGACCGCGCCCGACGACCTGGCTCCAGCTCGTCTCTCCCGCGCGCTGTTCATCCG 1174
 Db 306 -----LeuGlnTrpLeu----- 309
 QY 1175 AGCACAGAAATGAGACGACGCGCTAGCCTCTGCGACAGCTGTCCGACGAATGTCAG 1234
 Db 310 -----ArgArgIleGln 313
 QY 1235 CCGATATGTGTG---TAGGAGAGGTCAACCAAGCTGTGATCAATGTT 1279
 Db 314 AsnYrSerValMetAspIleCysAspYrAspGlnSerSerGlyArgTrp----- 330
 QY 1280 CATGACATCTTCTATCTCTCCCAATCAGAGGAGGACGAGAGCTGTGCTTCTCCGC 1339
 Db 330 ----- 330
 QY 1340 GCCAATGATGCAAGACGCGCTCTGCCAATTGTACAAAGTCAACGCGCTTTAAATCC 1399
 Db 331 ---AsnCysLeuValAlaAlaArgIleHisIleGlnMetSerThrThr----- 344
 QY 1400 CAGGCGTACGATTGGATGAGCGCTTCAGCCCGCGGGAAGATGA--- 1444
 Db 345 ---GlyTrpValGlyArgPheArgProSerGlnProHisPheThrLeuAspGly 361
 QY 1445 -----TTTAAG 1450
 Db 362 AsnSerPheYrLysIleIleSerAsnGlnLulGlyYrArgHisIleCysYrTrpHeGln 381
 QY 1451 TGCCCATTAAGAAAGAGATTGCTCTGACACGCGGTGAATGGAGATTGTCGCGAGCAC 1510

Db 382 IleAspLysLysAspCysThrPheIleThrLysGlyThrTrpGlnValIleGlyIleGlu 401
 QY 1511 GCGTCCAAAGATCTGGGTCAATGAGACGACCAAGCTGTGTACTTCCAGGCGACCAAGAC 1570
 Db 402 AlaLeu-----ThrSerAsp 406
 QY 1571 ACGCGCGTGAGACACACCTCTACGTGTGACGTATGAGGCGCGCGC----- 1618
 Db 407 -----YrLeuYrYrIleSerAsnGlnYrLysGlnMetProGlyGly 421
 QY 1619 ---GAGATCGTACCGCTTCACCGCCCGCGCTTCCCATCTGCTCCATGAGCCAGAAC 1675
 Db 422 ArgAsnLeuYrLysIleGlnLeuSerAspYrThrLysValIleHisLysSerCysGln 441
 QY 1676 TTCGACATGTTCTGTCAGCCACTACAGACGCTGACACGCGCGCTGCTGACGCTTAC 1735
 Db 442 LeuAsnProGlnArgCysGlnYrYrSerValSerPheSerLysGlnAlaLysYrYr 461
 QY 1736 AAGCTG-----AGCGGCGCC----- 1750
 Db 462 GlnLeuArgCysSerGlyProGlyLeuProLeuYrThrLeuHisSerSerValAsnAsp 481
 QY 1751 -----GACGACGACCCCTGACACAGCAGACCGCCGCTTGCGGCT 1789
 Db 482 LysGlyLeuArgValLeuGlnAspAsnSerAlaLeuAspLys----- 495
 QY 1790 AGCATGATGAGACGACCGACGCTGCCCCCGGATTAATGTCTCCAGACATTTCCATTTC 1849
 Db 496 ---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe 509
 QY 1850 CACACGCGCTCGGATGTGCGGCTC---TAGCGCATGATCTACAAAGCCCGACGCTTGAG 1906
 Db 510 IleIleLeuAsnGlnThrLysPheTrpYrGlnMetIleLeuProHis---PheAsp 528
 QY 1907 CCAGGGAAGACGACCCCGACGCTCTTGTATGTAGAGGCGCCCGACGTCAGCTGTG 1966
 Db 529 LysSerLysLysYrProLeuLeuAspValYrAlaGlyProCysSerGlnLysAla 548
 QY 1967 AATACTCTTCAAAAGCATCAAGTACTTGGCGCTCAAC-----ACACTGGCTTC 2017
 Db 549 AspThrValPhe-----ArgLeuAsnTrpAlaThrYrLeuAlaSer 562
 QY 2018 CTGGCGTACGGCGGTGTG---ATTGACGCGACGCGCTCTGTCACGAGGCGTTGG 2074
 Db 563 ThrGlnAsnIleIleValAlaSerPheAspGlyArgGlySerGlyYrLeuIleYrAspLys 582
 QY 2075 TTCGAAGGCGCCCTGAAACCAATGAGGCGGAGTGAATGAGACGAGCTGAGGCGC 2134
 Db 583 IleMetHisAlaIleAsnArgArgLeuGlyThrPheGlnValGlnAspGlnIleGlnAla 602
 QY 2135 CTG---CAGTTCGTGGCCGAGAGTATGAGCTTCATGACCTGAGCGGAGTTGCCAT 2191
 Db 603 AlaArgGlnPhe-----SerLysMetGlyPheValAspAsnLysArgIleAlaIleTrp 620
 QY 2192 GCGTGTCTACGCGGCGCTCTCTGCTCATGGGCGTAATCCAAACCCCGAGGTGTC 2251
 Db 621 GlyTrpSerYrGlyGlyYrValThrSerMetValLeuGlySerGlySerGlyValPhe 640
 QY 2252 AAGTGGCCATCGCGGCGGCGCGGCTGACCGCTGTGATGAGGCTACGACAGGATACACT 2311
 Db 641 LysCysGlyIleAlaValAlaProValSerArgTrpGlnYrYrAspSerValYrThr 660
 QY 2312 GAGCGCTACATGAGCGTC-----CTGAGAACAAACGACGAGCTATGAGCGCGGTTCC 2365
 Db 661 GlnArgYrMetGlyLeuArgProThrProGlnAspAsnLysArgIleAsnSerThr 680
 QY 2366 GTGGCGCTGACGCTGAGAGAGCTGCCAATGAGGCCCAACCGCTTGTATCTGTCAGGCG 2425
 Db 661 ValMetSerArgAlaGlnAsnPhe-----LysGlnValGlyYrLeuLeuIleHisGly 698
 QY 2426 TTCCTGACGAAACGTCACACTTTTTCACACAAACTCTGCTGTCGCAATGATTCGA 2485

```

Db 699 ThAlaAspAspAsnValHisPhgInglInserIaGlnIleSerLysAlaLeuValAsp 718
Oy 2486 GCAGGAAACCTTACACAGCTCCAGATCTACCCCAAGAGACACAGACTATTCGTCGCC 254
Db 719 ValGlyValAspPhgInGlnAlaMetIrrPtyrThrAspGlnAsnIleAlaSerSer 738
Oy 2546 GAGTGGGGCGAGCATATGAGAGTCAAGCTGGCTGCAGCTTCTACAGAGAA 2593
Db 739 ThrAlaHisGlnHisIleIrrThrHisMetSerHisPheIleLysGln 754

RESULT 3
US-08-230-491A-3
: Sequence 3, Application US/08230491A
: Patent No. 5587299
: GENERAL INFORMATION:
: APPLICANT: Retlgy, Wolfgang J.; Scanlan, Matthew J.;
: APPLICANT: Garlin-Chesa, Pilar; Old, Lloyd J.
: TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
: TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
: TITLE OF INVENTION: THEREOF
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: FELFE & LYNCH
: STREET: 805 THIRD AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: WORDPERFECT - ASC II
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/230,491A
: FILING DATE: 20-APRIL-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanson, No. 5587299man D.
: REGISTRATION NUMBER: 30,946
: REFERENCE/DOCKET NUMBER: LUD 330
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 688-9200
: TELEFAX: (212) 838-3884
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 766 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
US-08-230-491A-3

Alignment Scores:
Pred. No.: 1,04e-29 Length: 766
Score: 489.50 Matches: 201
Percent Similarity: 36.27% Conservative: 124
Best Local Similarity: 22.43% Mismatches: 330
Query Match: 10.21% Indels: 241
DE: 1 Gaps: 35

US-09-976-674-4 (1-2617) x US-08-230-491A-3 (1-766)
Oy 149 CTCATGTCGACCAAGGCGCCCGACGACTCCAGTTGTGCAAGAAGAGATGAGTCTGGG 208
Db 26 ValLeuLeuAsnLysGlyI- - - - - ThrAspAspAlaThr 36
Oy 209 CCCACCTCCACCGGCTCTACTACTGCGGAATGCCATATGGACGCCAGAGAACTCCCTC 268
Db 37 AlaAspSerArgLysThrIrrThrLeu- - - - - 45
Oy 269 CTCCTACTGAGATTCCCAAGAGAGTCCGGAAGAGAGGCTGTCTGCTCCTGCTCGAGAG 328
Db 46 - - - - - ThrAspTyrLeuLysAsnThrTyrIrrGluLeuLysLeuTyrSerLeuArgTrrp- - - 62

```

OY	329	CAGATGCGGATCACTTTCCAGGGCAGGCCGCCACCAATGGGGCTACTACTCGGGAGGAGGAG	388
Db	63	-----IIeserAPhISgluTYrLYsglnIuAsn	75
OY	389	CTGGTG-----AGGAGCGGAAAGCCTG	412
Db	76	IIleuValrPheAsnIaIaGluTYrGlyAsnSerSerValrPheIuGlnIuAsnSerThrPhe	95
OY	413	GGGGCTTTCGGCATCACCTCTTACAGCTTCCAG--AGCGAGATGGCCCTTCTCTTC	469
Db	96	AspIuPheGlyHisSerIIeAsnAspTYrSerIIeserProAspGlyClnPheIIeIu	115
OY	470	CAGCGCAGCAACACCCTTCCACTGTCGGCGAGCGGGCGGAACAAGCGCTCATGCTGCC	529
Db	116	IeugluTYrAsnTYrValLYsglnTYrPaGlnHisSerTYrThrAlaSerTYrAspIIeYr	135
OY	530	CCATGAACCGCGTGGAAATCAAGACCCAGCTGTACGGCCCCGGATGAGCCCAATTC	589
Db	136	AspIeAsnIuysaTGlnIuIeThrIuGlnIuArgIIeProAsnAsnThrGlnTYrAla	155
OY	590	TGCCCTGGCCGACCTGCTTC--TTCTCCATCATTAACAGCAGCCTGTGGGTGCC	646
Db	156	ThrTYrSerProValGlyHisLYsLeuAlaTYrValTYrPaAsnAspIIeYrValLYs	175
OY	647	AACATCGAGACAGCGCAGAGCGGGCGGTGACCTTCTGGCCACCAAGTTTATCCATTC	706
Db	176	IIeugluProAsnIeuproSerTYrArgIIeThrTYr-----ThrGlyLYsglnIuAspIIe	193
OY	707	CTGGATGACCCCAAGTGTGGGGGTGTGGCCACTTGTGCATACAGAGAG--TTCCAGC	763
Db	194	IIeYrAsn-----GlyIIeThrAspTYrValTYrGlnIuGlnIuAlaPheSer	209
OY	764	CGCTTACCTGGGTACTGTGTGGTCCGCCACAGCCTCTGGGAAGTTCAAGGGCCTCAG	823
Db	210	AlaTYrSerAlaIeUtrTYrTYrSerProAsnGlyThrPhe-----	222
OY	824	ACGCTGGCAATCTGTATGAGAAATCGATGAGTCCGAGGTGAGAGCTATTCAC-----	877
Db	223	-----LeuAlaTYrAlaGlnIuPheAsnAspThrGluValProIeIuIeIuTYrSer	239
OY	878	---GTCCTCTCTCTGGGTGCTAGAAAGAAAGAACAGCAGCATCTTACCGGACCCGAGCA	934
Db	240	PheTYrSerAspGlnSerIeuglnTYrProLYsThrValaIArgValaIProTYrProLYsAla	259
OY	935	GGCAGCAGAATCCCAAGATGTGCTTGAACCTGGCTGAGTTCAGACTGACAGCCAGCGC	994
Db	260	GlyAlaValaAsnProThrVal-----LYsPhePheValaIValAsnThrAspSerIeUser	277
OY	995	AAGATCTCTTCGACCCAGGAGAAAGAGACTGTGTACACCTTAGCTCGCTGTCCCAAG	1054
Db	278	SerValaIThrAsnAlaThrSerIIeGlnIeThrAlaIProAlaSerMetIeUleuIIeLYsP	297
OY	1055	GTGGAGTACATCGCCAGGCGCGGTGTGACCCCGGAGTGCAGAAATPACGCGGGCCATGTC	1114
Db	298	---HisTYrIeUcyAspValaIThrTYrAlaIThrGlnIuArgIIeSer-----	312
OY	1115	CTGGACGGGCCCCAGCAGTGTGCTCAGACTGTCTCTCCGCCGGCCTGTATCCCG	1174
Db	313	-----IeuglnTYrIeU-----	316
OY	1175	AGCACAGAAATGAGAGCAGCGGCTTAGCTCTGCCAGAGCTGTCCAGAAATGTCCAG	1234
Db	317	-----ArgArgIIeIu-----	320
OY	1235	CCGATGTGGTG-----TAGAGAGAGGTCAACAGCTGTGATCAATGTT	1279
Db	321	AsnTYrSerValaIuMetAspIIeCYsAspTYrAspGlnSerSerGlyaIArgTYr-----	337
OY	1280	CATGACATCTTATCCCTTCCCAATAGAGAGGAGAGCAGACTCTGCTTCTCCGC	1339
Db	337	-----	337
OY	1340	GCCAAATGATCAAGACCGGCTTCTGCCATTTGTACAAAGTCAACGCGGCTTTAAATTC	1399

```

Db 338 -----AsnCySLeuValAlaArgGlnHisIleGluMetSerThrThr----- 351
QY 1400 CAGGGCTACGATTGGATGAGCCCTTCAGCCCGGGGAGATGAA----- 1444
Db 352 -----GlyTrpValGlyArgPheArgProSerGluProHisPheThrLeuAspGly 368
QY 1445 -----TTTAAG 1450
Db 369 AsnSerPheThrIleIleSerAsnGluGluGlyTrpArgHisIleCysTyrPheGln 388
QY 1451 TGCCCATTAAGGAAGAGATTGCTGTGACGCGTGAAATGGAGGATTGGCAGGCAC 1510
Db 389 IleAspLysAspCysThrPheIleThrLysGlyThrTrpGluValIleGlyIleGlu 408
QY 1511 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGGTACTTCCAGGGCAGCAAGAC 1570
Db 409 AlaLeu-----ThrSerAsp 413
QY 1571 ACGCCGCTGGAGCACACCTCTACGTGCTAGTATGAGCGCGCGG----- 1618
Db 414 -----TyrLeuTyrTrpIleSerAsnGluTyrLysGlyMetProGlyGly 428
QY 1619 ---GAGATCGTACGCTCACACACGCGCGCTTCTCCATAGCTGCTCATGAGCCAGAAC 1675
Db 429 ArgAsnLeuTyrLysIleGlnLeuSerAspTyrThrLysValThrCysLeuSerCysGlu 448
QY 1676 TTCGACATGTTCTGTCAGCCACTACAGACGCTGAGACGCGCGCTGCTGACGTTAC 1735
Db 449 LeuAsnProGluArgCysGlnTyrTyrSerValSerPheSerLysGluAlaLysTyrTyr 468
QY 1736 AAGCTG-----ACGCGGCC----- 1750
Db 469 GlnLeuArgCysSerGlyProGlyLeuProLeuTyrThrIleuHisSerSerValAsnAsp 488
QY 1751 -----GACGACGACCCCTGCACAGACGCGCGCTTCTGGGCT 1789
Db 489 LysGlyLeuArgValLeuGlnAspAsnSerAlaLeuAspLys----- 502
QY 1790 AGCATGATGAGGAGCAGCAGCTGCCCGGATATGTTCTCTCCAGAGATTTTCATTTC 1849
Db 503 ---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe 516
QY 1850 CACACGGCTGCGATGTCGGCTC---TAGCGCATGATGTCAGAGCCGCGCTTGAG 1906
Db 517 IleIleLeuAsnGlnTyrLysPheTyrTrpGlnMetIleLeuProHis---PheAsp 535
QY 1907 CCAGGAGAAGAGACACCCCGCTCTTTGATATGAGAGCCCGCAGGTGACGTGCTG 1966
Db 536 LysSerLysLysTyrProLeuLeuAspValTyrIleGlyProCysSerGlnLysAla 555
QY 1967 AATACCTCTTCAAAAGCATCAAGTACTTGGCGCTCAAC-----ACACTGCGCTCC 2017
Db 556 AspThrValPhe-----ArgLeuAsnTrpAlaThrTyrLeuAlaSer 569
QY 2018 CTGGGCTACGCGCGTGTG---ATTGAGGAGGAGGCGCTCTGTCAGCGGCGTTCGG 2074
Db 570 ThrGlnAsnIleIleValAlaSerPheAspGlyArgLysGlyTyrGlnLysAspLys 589
QY 2075 TTCGAAGGGCGCTCAAAAACCAATGAGGCGAGGTGAGATCGAGAGCAGGTGAGGCG 2134
Db 590 IleMetHisAlaIleAsnArgArgLeuGlyThrPheGluValGluAspGlnIleGluAla 609
QY 2135 CTG---CAGTTCGTGGCGGAGAGATGAGCTTTCATGACCTGAGCCGAGTTGCCATCAT 2191
Db 610 AlaArgGlnPhe-----SerLysMetGlyPheValAspAsnLysArgIleAlaIleTrp 627
QY 2192 GGCAGTCTACGAGGCGCTCTCTGCTCATGGGCGTAAATCCACAAACCCAGGTGCTC 2251
Db 628 GlyTrpSerTyrGlyGlyTyrValThrSerMetValLeuGlySerGlySerGlyValPhe 647
QY 2252 AAGGTGGCATGCGCGGTGCCCGGTACCGCTCTGATGGCTTACGACACAGGATCACT 2311

```

```

Db 648 LysCysGlyIleAlaValAlaProValSerArgTrpGluTyrTyrAspSerValTyrThr 667
QY 2312 GAGCGCTACATGAGACGTC-----CTGAGAACACACGACGCGCTATGAGCGGTTCC 2365
Db 668 GlnArgTyrMetGlyLeuProThrProGlnAspAsnLeuAspHisTyrArgAsnSerThr 687
QY 2366 GTGGCCCTGCGACGTGAGAAAGCTGCCCAATGAGGCCCAACGCGTGTTCATCTTCACGCG 2425
Db 688 ValMetSerArgAlaGluAsnPhe-----LysGlnValGluTyrLeuLeuIleHisGly 705
QY 2426 TTCCTGACGAAACAGTCGACCTTTTTCACACAAACTCTCTGCTGCCAATGATCCGA 2485
Db 706 ThrAlaAspAspAsnValHisPheGlnGlnSerAlaGlnIleSerLysAlaLeuValAsp 725
QY 2486 GCAGGAACCTTACCAAGCTCCAGATCTACCCCAACGAGACACACAGTATTCGTCGCC 2545
Db 726 ValGlyValAspPheGlnAlaMetTrpTyrThrAspGluAspHisGlyIleAlaSerSer 745
QY 2546 GAGTCGGCGGACGACATGAAATGACGTCGCTGCTGCTTCTACAGAA 2593
Db 746 ThrAlaHisGlnHisIleTyrThrHisMetSerHisPheIleLysGln 761

```

```

RESULT 4
US-08-619-280A-3
; Sequence 3, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Retlig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Feltz & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-619-280A-3

```

```

Alignment Scores:
Pred. No.: 1,04e-29 Length: 766
Score: 489.50 Matches: 201
Percent Similarity: 36.27% Conservative: 124
Best Local Similarity: 22.43% Mismatches: 330
Query Match: 10.21% Indels: 241
DB: 1 Gaps: 35

```

US-09-976-674-4 (1-2617) x US-08-619-280A-3 (1-766)

QY 149 CTCATTGTCAACAGGCGCCACGACCTCCAGTTGTGACGAAGACGATGAGTGG 208
: : : : :
Db 26 ValLeuLeuasnLysGly-----ThrAspAlaThr 36
QY 209 CCCCACTCCACCGCTCTACTACTGGAATGCCATATGACGCCGAGACACTCCCTC 268
: : : : :
Db 37 AlaAspSerAlaGlySthrTyThrLeu----- 45
QY 269 CTCACCTGATGATTCACCAAGAGGCTGCTGCTCTCTCTCTGAG 328
: : : : :
Db 46 -----ThrsPylLeuLysAsnThrTyTrArgLeuLysLeuTySerLeuAlaTrp-- 62
QY 329 CAGATGCTGATCATTTCCAGGCGCACGCCACCATGGGCTACTCTCGGAGAGAG 388
: : : : :
Db 63 -----IleSerAspHisGluTyThrLeuTyTrLysGlnGluAsnAsn 75
QY 389 CTGCTG-----AGGAGCGGAAAGCGCTG 412
: : : : :
Db 76 IleLeuValPheAsnAlaGluTyTrGlyAsnSerSerValPheLeuGluAsnSerThrPhe 95
QY 413 GGGGCTTCGGCATCAGCTCTGACACTCCAC--AGCGAGATGGCTCTCTCTCTC 469
: : : : :
Db 96 AspGluPheGlyHisSerIleAsnAspTySerIleSerProAspGlyGlnPheIleu 115
QY 470 CAGGCCAGCAACAGCTCTTCCACTGCGGACGCGGCAAGAACGGCTTCATGCTGCC 529
: : : : :
Db 116 LeuGluTyTrAsnTyTrValLysGlnTrpArgHisSerTyThrAlaSerTyTrAspIleTyTr 135
QY 530 CCTATGAACCCGCTGGAAATCAGACCCAGTCTCAGGCGCCCGGATGAGCCCAAAATC 589
: : : : :
Db 136 AspLeuAsnLysArgGlnLeuIleThrGluGluArgIleProAsnAsnThrGlnTrpVal 155
QY 590 TGCCCTCCGACCCCTGCTCTC--TTCCTCTCATCATATACAGACGCTGGTGGGCC 646
: : : : :
Db 156 ThrTrpSerProValGlyHisLysLeuAlaTyTrValTrpAsnAspIleTyTrValLys 175
QY 647 AACATCGACAGCGGAGGAGCGCGCTGACCTTCGCCACCAAGGTTATCCAAATGTC 706
: : : : :
Db 176 IleGluProAsnLeuProSerTyTrArgIleThrTrp-----ThrGlyLysGluAspIle 193
QY 707 CTGATGACCCCAAGTGTGGGGTGTGGCCACTCTCATACAGGAAG--TTGAC 763
: : : : :
Db 194 IleTyTrAsn-----GlyIleThrAspTrpValTyTrGluGluValPheSer 209
QY 764 CGCTTCACTGGTACTGTGTGTCGCCACAGCCTCGGAGGTTGAGAGGCTCAG 823
: : : : :
Db 210 AlaTySerAlaLeuTyTrPTrSerProAsnGlyThrPhe----- 222
QY 824 AGCCTGCGAACTCTGATVGAGAAAGTCGATGATCGAGGTGAGATCATTCAC----- 877
: : : : :
Db 223 -----LeuAlaTyTrAlaGlnPheAsnAspThrLysValProLeuIleGluTySer 239
QY 878 ---GTCCCTCTCTCGGCTGAGAAAGAAAGACGACCTGATCGGTACCCAGACA 934
: : : : :
Db 240 PheTySerAspGluSerLeuGlnTyTrProLysThrValArgValProTyTrProLysAla 259
QY 935 GGCAGCAAGATCCCAAGATTTGCTTGAACATGCGTGAATTCAGACAGGCG 994
: : : : :
Db 260 GlyAlaValAsnProThrVal-----LysPhePheValValAsnThrAspSerLeuSer 277
QY 995 AAGATGCTCGACCCAGAGAGAGAGCTGTCAGCCCTTCAGCTGCTGCCGAG 1054
: : : : :
Db 278 SerValThrSerAlaThrSerIleGlnIleThrAlaProAlaSerMetLeuIleGlyAsp 297
QY 1055 GTGAGATACATGCCAGGCGGGGTGACCCGGATGAGCAATACGCTGGGCATGTTTC 1114
: : : : :
Db 298 ---HisTyTrLeuLysAspValThrTrpAlaThrGlnGluArgIleSer----- 312
QY 1115 CTGAGCCGCGCCAGGAGGCTGCTCAGCTGCTCTCCCGCGCGCTTCATCCCG 1174
: : : : :
Db 313 -----LeuGlnTrpLeu----- 316

QY 1175 AGCAGAGAAATGAGAGACAGCGGCTAGCCTCTGACAGAGCTGCCCAAGATGTCAC 1234
: : : : :
Db 317 -----ArgArgIleGln 320
QY 1235 CCGTATGTGCTG-----TACGAGAGGTCAACCAAGTGTGATCAATGTT 1279
: : : : :
Db 321 AsnTySerValMetAspIleCysAspTyTrAspLysSerGlyArgTrp----- 337
QY 1280 CATACATCTTCTATCCCTTCCCAATCCAGAGGAGAGAGAGCTGCTTCTCCGC 1339
: : : : :
Db 337 ----- 337
QY 1340 GCCAATGATGACAGACCGGCTTCTGCCATTGTACAAAGTACACCGCGTTTAAATCC 1399
: : : : :
Db 338 -----AsnGlyLeuValAlaArgGlnHisIleGluMetSerThrTrp----- 351
QY 1400 CAGGCTACGATTTGAGTGAAGCCCTTCAGCCCGGAGAGATGAA----- 1444
: : : : :
Db 352 -----GlyTrpValGlyArgPheArgProSerGluProHisPheThrLeuAspGly 368
QY 1445 -----TTTAAAG 1450
: : : : :
Db 369 AsnSerPheTyTrLysIleIleSerAsnGluGluGlyTyTrArgHisIleCysTyTrPheGln 388
QY 1451 TGCCCATTAAGAGAGAGATTTGCTGACACGCGTGAATGAGAGGTTTGGCGAGGCAC 1510
: : : : :
Db 389 IleAspLysLysAspGlySthrPheIleThrLysGlyThrTrpGluValIleGlyIleGlu 408
QY 1511 GCGTCCAGATCTGGGTCAATGAGAGACCAAGCTGATCTTCCAGGCGACCAAGGAC 1570
: : : : :
Db 409 AlaLeu-----ThrSerAsp 413
QY 1571 ACGCGCTGAGACACACCTCTACGTGACATGAGGCGCGCGC----- 1618
: : : : :
Db 414 -----TyLeuTyTrLysIleSerAsnLysGlyMetProGlyGly 428
QY 1619 ---GAGATCGTAGCCCTCACACGCGCGGCTTCCCATGAGCTTCATGACAGCAAGAC 1675
: : : : :
Db 429 ArgAsnLeuTyTrLysIleGlnLeuSerAspTyTrThrLysValThrCysLeuSerCysGlu 448
QY 1676 TTGCACATGTTTGTGACACCACTACAGAGCGTGAACGCGCGCTCGTCACAGTCTAC 1735
: : : : :
Db 449 LeuAsnProGluArgCysGlnTyTrTySerValSerPheSerLysGluAlaLysTyTrTyTr 468
QY 1736 AAGCTG-----AGCGGCGCC----- 1750
: : : : :
Db 469 GlnLeuArgCysSerGlyProGlyLeuProLeuTyTrThrLeuHisSerSerValAsnAsp 488
QY 1751 -----GACGAGACCCCTGACAGAGCGCCGCTTCTGGGCT 1789
: : : : :
Db 489 LysGlyLeuArgValLeuGluAspAsnSerAlaLeuAspLys----- 502
QY 1790 AGCATGTGAGAGCGACGACCTGCCCCCGGATTAATGTTCTCTCAGAGATCTTCATTTC 1849
: : : : :
Db 503 ---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe 516
QY 1850 CACACGCGCTCGGATGCGCGCTC--TACGCGATGATTAACAAGCCCGCGCTTGAC 1906
: : : : :
Db 517 IleIleLeuAsnGlnThrLysPheTrpTyTrGlnMetIleLeuProProHis---PheAsp 535
QY 1907 CCAGGGAAGAGACCCACCGCTCTTGTATATGAGGCCCCCAGGATGACGTGGTG 1966
: : : : :
Db 536 LysSerLysLysTyTrProLeuLeuLeuAspValTyTrAlaGlyProCysSerGlnLysAla 555
QY 1967 AATAACTCTTCAAAAGCATCAAGTACTTGGCTCAAC-----ACACTGGCGCTC 2017
: : : : :
Db 556 AspThrValPhe-----ArgLeuAsnTrpAlaThrTyTrLeuAlaSer 569
QY 2018 CTGAGCTACCGCTGGTGTG--ATTGACGAGGAGGCGCTCTGACAGAGGCGCTTCGG 2074
: : : : :
Db 570 ThrGluAsnIleIleValAlaSerPheAspGlyArgGlySerGlyTyTrGlnGlyAspLys 589

QY 2075 TTCGAGGCGCCCTGAAAAACCAATGGCCAGGTGAGATGAGACAGGAGGCG 2134
 Db 590 ILEMETHISALAIASARGARGLEUCLYTHRPHGELVALGVALSPGILLLEGLNLA 609
 QY 2135 CTG---CAGTTCCGGCCGAGAACTATGCGTTTCATCCAGCTGACCCGAGTTGCCATCCAT 2191
 Db 610 AIAARGINPHE-----SERLYSMETGLYPHEVALASPAISLYSARGILEALILETRP 627
 QY 2192 GCGTGGTCCTACGCGGGGCTCTCGCTCATGGGGCTTCAATCCACAAGCCCGAGGTCTC 2251
 Db 628 GLIYRPERLYTGLYGLYTYVALTHRSETMETVALLEUGLYSERGLYVALPHE 647
 QY 2252 AAGGTGCGCATCGCGGGTCCCGCGTCAACCTGTGATGGCTCAGACACAGGTACACT 2311
 Db 648 LYSGLYSGLYLEALVALAIALPROVALSERARGTRPGIUTYTYRASPSEVALIYLYTHR 667
 QY 2312 GAGCGCTACATGACGTC-----CTGAGAACAAACGACGAGCGCTATGAGCGGGTTC 2365
 Db 668 GLIARGLYMETGLYLEUPROTHRPROGLIUSPASNLEUASPHISLYTYRARGASERTHR 687
 QY 2366 GTGGCCCTGCACGCTGGAGAACTGCCCATGAGCCCAACCCCTGTATCTCCACGCG 2425
 Db 688 VALMETSERARGALAGIUSAPNHE-----LYSGIIVALGLIUTYRLEULEIHISGLY 705
 QY 2426 TTCCTGGACGMAAAGCTGCACTTTTCCACAAACTCTCTGTCGCCAAGTATCCGA 2485
 Db 706 THRILASPSAPSNVALHISPHGELINGINLSERLISERLYSALILEUVALASP 725
 QY 2486 GCAGGGAACCTTACCAAGCTCCAGATCTACCCCAACGAGACAGAGATATTCGTCGCC 2545
 Db 726 VALGLIYVALSPRHEGINALAMETRPTLYTHRASPGLIUSAPNHSGLYIILALASER 745
 QY 2546 GAGTCGGGCGGACGACTATGANGTCAAGCTTGCTGCACTTCCACAGGA 2593
 Db 746 THRILASGLINHSILIEYRTHRHSIMETSERHISPHLEIYSGIN 761
 RESULT 5
 US-08-940-391-3
 ; Sequence 3, Application US/08940391
 ; Patient No. 5965373
 ; GENERAL INFORMATION:
 ; APPLICANT: Zimmermann, Rainer; Park, John E.;
 ; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
 ; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
 ; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Feltz & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: WordPerfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/940,391
 ; FILING DATE: 01-OCT-1997
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/619,280
 ; FILING DATE: 18-MARCH-1996
 ; APPLICATION NUMBER: 08/230,491
 ; FILING DATE: 20-APRIL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5965373man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5330.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 766 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 US-08-940-391-3
 Alignment Scores:
 Score: 1.04e-29 Length: 766
 Percent Similarity: 489.50 Matches: 201
 Best Local Similarity: 36.27% Conservative: 124
 Query Match: 22.43% Mismatches: 330
 DB: 10.21% Indels: 241
 Gaps: 35
 US-09-976-674-4 (1-2617) x US-08-940-391-3 (1-766)
 QY 149 CTCATTTGCACAGAGCGCCCGCCAGACTTCCAGTTTGTGCAGAAAGAGATGAGTGGG 208
 Db 26 VALLEUENASNLYSGLY-----THASPSPALATHR 36
 QY 209 CCCCACTCCACCGCCTCTACTACTGGGAATGCCATATGCGAGCGAGAACTCCCTC 268
 Db 37 ALASPSERARGLYSTHRTYRTHLEU----- 45
 QY 269 CTCATCTGAGATCCCAAGAGGCTCCGAAAGAGCTCTGCTCTCTCGAAG 328
 Db 46 -----THASPTYLEULYASNTHTYRARGLEULYSLEUTYRSELEUADTRP--- 62
 QY 329 CAGATGCTGATCATTTCCAGGCCACGCCCATGAGGAGGAGAGAG 388
 Db 63 -----ILESERPHISGLIUTYRLEUTYRGLINLUSPASN 75
 QY 389 CTGCTG-----AGGACGCGMAAGCCCTG 412
 Db 76 ILEUVALPHEASNALAGIUTYRGLYASNSERVALPHEUGLIUSASERTHRPE 95
 QY 413 GGGCTTTCGGCATCAGCTCTAGACTCCAC---ACGAGAGTGGCTTCTCTCTC 469
 Db 96 ASPLUPRHEGLYHISSEIILASNPYRISERILESERPROASPGIYGLNPHLEU 115
 QY 470 CAGGCGACACAGCGCTTCCAGCTGTCGCGCGGCGGCAAGAACGCTCATGTCGTC 529
 Db 116 LEUGIUTYRASNLYVALYSGIINTPRAGHISERTYRTHRALASERTYRASPILYR 135
 QY 530 CCTATGAACCGCTGGAATCAAGACCCAGCTCAGGCGCGGATGAGCCCAAAATC 589
 Db 136 ASPLEUASNLYSARGINLEULETHRGILUAGTLEPROASNTHRGINTPRVAL 155
 QY 590 TGCCCTGCCAGCCTGCTTC---TTCCTTCATCAATTAACGAGCCTTGCGTCC 646
 Db 156 THRTPRSERPROVALGLYHISLYSLEUALETYRVALTPRASNASNPILLETYRVALYS 175
 QY 647 AACATCGAGACAGGAGAGCGGCGGCTGACCTTCCACCAAGTTTATCAATGTC 706
 Db 176 ILEGLUPROASNLEUPROSERTYRARGILETHRTP-----THRGILYSGIUSPIL 193
 QY 707 CTGATGACCCCAAGTCTGCGGTGTGGCCACTTCGTGATACAGAGAGAG---TTCGAC 763
 Db 194 ILETYRASN-----GLYILERHASPRTIPVALTYRGLIUGIUGIUALPHESER 209
 QY 764 CGCTTCACTGGGTACTGCTGTGCTGCCACAGCCTTGGGAAGTTCAAGAGGCGCTCAAG 823
 Db 210 ALATYRSEALALEUTIPTRIPSERPROASNLYTHRPE----- 222
 QY 824 ACGTGCGAATCTGTATGAGGAAGTGAATGATGAGTCCGAGGATTCAC----- 877
 Db 223 -----LEUALATYRALAGIINPHEASNAPNTHRGIUALPROLEIILEGLIYR 239
 QY 878 ---GTCCCTCTCTGCGCTAGAGAAAGAGAGAGAGCTGTATCGGTACCCAGAGACA 934
 Db 240 PHETYRSEIASPGLUSERLEUGINTYRPROLYSTHRTVALARGVALPROTYRPROLYSALA 259

[illegible][illegible]

Alignment Scores:


```

Db 556 AspThrValPhe-----ArGleuAsnTrpAlaThrTyrLeuAlaSer 569
QY 2018 CTGGGGCTACCGCCGTTGTG---ATTGACGCGAGGGCTCCTGTCAGCGGGCTCGG 2074
Db 570 ThGluAsnIleIleValAlaSerPheAspGlyArgGlySerGlyTyrGlnGlyAspLys 589
QY 2075 TTCGAAAGGGCCCTGAAACCAATGAGGCGAGTGCAGATCGAGACCAAGTGGAGGC 2134
Db 590 IleMetHisAlaIleAsnArgArgLeuGlyThrPheGluValGluAspGlnIleGluAla 609
QY 2135 CTG---CAGTTCTGGGCCCGAGATAGGCTTCATCGACTGAGCCGAGTTGCCATCAT 2191
Db 610 AlaArgGlnPhe-----SerLysMetGlyPheValAspAsnLysArgIleAlaIleTrp 627
QY 2192 GGCCTGCTACGAGGGGCTTCCTCGCTCATGGGCTAAATCCACAGGCCCGGCTGTC 2251
Db 628 G1YTrpSerTyrGlyGlyTyrValThrSerMetValLeuGlySerGlySerGlyValAlaPhe 647
QY 2252 AAGGTGGCATCGCGGGTCCCGCCGTCACCGTGTGATGGCTTCAGACAGAGGCTACACT 2311
Db 648 LysCysGlyIleAlaValAlaProValSerArgTrpGlyTyrArgSerValTyrThr 667
QY 2312 GAGCGCTACATGACGTC-----CCTGAGACACACACAGCGGCTATGAGCGGGTTCC 2365
Db 668 GluArgTyrMetGlyLeuProThrProGluAspAsnLeuAspHisTyrArgAsnSerThr 687
QY 2366 GTGGCCCTGCAGCGTGGAGAGCTGCCAATGAGCCCAACCGCTTCATTCCTGCACGGC 2425
Db 688 ValMetSerArgAlaGluAsnPhe-----LysGlnValGluTyrLeuIleHisGly 705
QY 2426 TTCCTGACGACAAACGTGCACTTTTTCACACAAACTCTCGTCCCACTGATCCGA 2485
Db 706 ThrAlaAspAsnValHisPheGlnGlnSerAlaGlnIleSerLysAlaLeuValAsp 725
QY 2486 GCAGGGAACCTTACCGAGCTCCAGATCTACCCACAGAGACACATATTCGTCGCC 2545
Db 726 ValGlyValAspPheGlnAlaMetLeuTrpTyrThrAspGluAspHisGlyIleAlaSer 745
QY 2546 GAGTCGGGCGACACTATGAAGTCAGCTGTCGACTTTCACAGGA 2593
Db 746 ThrAlaHisGlnHisIleTyrThrHisMetSerHisPheIleLysGln 761

RESULT 7
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 555X
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/833,211
; FILING DATE: 06-FEB-1992
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US93-07923-3

Alignment Scores:
Pred. No.: 3.61e-29 Length: 755
Score: 482.50 Matches: 194
Percent Similarity: 36.47% Conservative: 116
Best Local Similarity: 22.82% Mismatches: 321
Query Match: 10.06% Indels: 219
DB: 5 Gaps: 33

US-09-976-674-4 (1-2617) x PCT-US93-07923-3 (1-755)
QY 287 AAGAAAGTCCGGAAGAGAGCGCTGCTGCTGCTGTAAGAGCAGATGTCATATTC 346
Db 39 LysAsnThrTyrAlaGluLeuLysLeuTyrSerLeuAlaTrp----- 51
QY 347 CAGGCCAGCCGCCACATGGGGTCTACTCTGGGAGAGAGACTGTCG----- 394
Db 52 ---IleSerAspHisGlyTyrLeuTyrLysGlnLysAsnIleLeuValPheAsnAla 70
QY 395 -----AGGAGCGGAAAGCCCTGGGGGCTTCGGCATACC 430
Db 71 GluTyrGlyAsnSerSerValPheLeuGluAsnSerThrPheAspGluPheGlyHisSer 90
QY 431 TCCATACGACTTCCAC---AGCAGAGTGGCCTTCCCTTCCAGGCCACACAGCCTC 487
Db 91 IleAsnAspTyrSerIleSerProAspGlyGlnPheIleLeuGluTyrAsnTyrVal 110
QY 488 TTCACATGTGGCGACGGCGGCAGACAGCGCTTCATGCTGCCCTATGAACCGCTGAA 547
Db 111 LysGlnTrpArgHisSerTyrThrAlaSerTyrAspLeuIleValAsnLysArgGln 130
QY 548 ATCAAGACCCAGTCTGAGGCGCCCGGATGAGCCCAAAATCTGCCGACCCCTGCC 607
Db 131 LeuIleThrGluGluArgIleProAsnAsnThrGlnTrpValThrTrpSerProValGly 150
QY 608 TTC---TTCCTCTTCATCAATACAGCGACCTGGGTGGCCCAATCGACAGCGGAG 664
Db 151 HisLysLeuAlaTyrValIleTrpAsnAsnAspIleTyrValLysIleGluProAsnLeuPro 170
QY 665 GAGCGGCGCTGACCTTCTGCCACCAAGTTTATCCAAATGCTCGATGACCCCAAGTCT 724
Db 171 SerTyrArgIleThrTrp-----ThrGlyLysGluAspIleIleTyrAsn----- 185
QY 725 GCGGGTGTGGCCACCTTCGTCATACAGGAAGAG---TTCAGCCGCTTCATCGGTACTGG 781
Db 186 ---GlyIleThrAspTrpValTyrGluGluGluValPheSerAlaTyrSerAlaLeuTrp 204
QY 782 TGGTGGCCCAAGCGCTCTCTGGAGAGTTTCAGAGGGCTCAAGACGCTGCAGAACTCTAT 841
Db 205 TrpSerProAsnGlyThrPhe-----LeuAlaTyr 214
QY 842 GAGCAAGTCATGATGTCGAGTGGAGTCAATTCAC-----GTCCCTCTCTGCG 892
Db 215 AlaGlnPheAsnAspThrGluValProLeuIleGluTyrSerPheTyrSerAspGluSer 234
QY 893 CTGAAGAAAGGAAGAGCGACTGTATCGGTACCCCGAGACAGGACAGCAAGATCCCAAG 952
Db 235 LeuGlnTyrProLysThrValArgValProTyrProLysAlaGlyAlaValAsnProThr 254
```

OY	953	ATTCGCTTGAAC	TGGCGGTAGTTCCAGAGCAGACGACGAGCAAGACAATGCTGCACACCAG	1012
Dd	255	Val-----	-LysPhePheValValAsnThrAspSerLeuSerSerValThrasnAlaThr	272
OY	1013	GAGAAGAGCGTG	GGGAGGCCCTTCACGTCGCTGTTCCGAAGTAGAGTAATCATGCCAGG	1072
Dd	273	SerIleGlnIle	ThrAlaProAlaSerMetLeuIleGlyAsp---HisTryLeucylasp	291
OY	1073	GCCGGGTGGAC	CGGGATGGCAAAATACGCCGTGGCCATGTCTTGACCGGCCAACGAG	1132
Dd	292	ValThrTrpAla	ThrGlnArgIleSer-----LeuGln	303
OY	1133	TGGCTCCAGCT	CGTCTCTCCCCCGGCCGCTTCATCCAGACAGAGATGAGAG	1192
Dd	304	Tripleu	-----	305
OY	1193	CAGCGGTAGCT	CTTCGCCAGAGCTGTCCCAGAAATGTCCAGACCGATTGTGTTG-	1246
Dd	306	-----	-----ArgArgIleGlnAspTrySerValMetasp	315
OY	1247	-----	-TACGAGAGCTCACCAACGTCGTATCAATGTCATGACATCTTCATCC	1297
Dd	316	IleCysAspTry	AspGluSerSerGlyAspTrp------	326
OY	1298	TTCCCCCAAT	CAGAGGAGAGACGAGACTGTCTTCTCCGCGCCATAAATGCAAGACC	1357
Dd	327	-----	-----AsnCysIleuVal	330
OY	1358	GGCTTCCTGC	ATTGTGTCAAAGTACACCGCCGTTTTAAATCCAGAGGTACGATTGGAGT	1417
Dd	331	AlaArgGln	HisIleGluMetSerThr-----GlyTrpVal	343
OY	1418	GAGCGCTTCA	CGCCCCCGGAGAAATGA--	1444
Dd	344	GlyArgPheArg	ProSerGluProHisPheThrLeuAspGlyAsnSerPheTryIle	363
OY	1445	-----	-----TTTAAGTCCCATTAAGCAAGAG	1468
Dd	364	IleSerAsnGlu	GluGlyTrgArgHisIleCysTryPheGlnIleAspLysAspCys	383
OY	1469	ATTCGCTGAC	CAGCGGTGAATGAGAGCTTTGGCAGGACGCGTCCAAGATCTGGGTC	1528
Dd	384	ThrPheIle	ThrLysGlyThrTrpGluValIleGlyIleGluAlaLeu-----	399
OY	1529	AATGAGAGAC	CAAGCTGTACTTCCAGGGCACCAAGACACGCGCTGAGCACACAC	1588
Dd	400	-----	-----ThrSerAsp-----Tyr	403
OY	1589	CTTACGCGGT	CACCTATGAGGGGCGCGC-----GAGATCGTACGCTC	1633
Dd	404	LeuArgTryIle	SerAsnGlnTryLysGlyMetProGlyGlyArgAsnIleuTryIle	423
OY	1634	ACCACGCCG	CGCTTCCCATACCTGCTCCATCAGGACCAAACTTGACATGTCCTCAGC	1693
Dd	424	GlnIleuSer	AspTryThrLysValThrCysLeuSerCysGlnLeuAsnProGluArgCys	443
OY	1694	CACATACAGAC	GTGAGCACGCGCGCTGCGTCCACGTTCACAGCTG-----AGCGGC	1747
Dd	444	GlnTryTrpSer	ValSerPheSerLysGluAlaLysTryGlnLeuArgCysSerGly	463
OY	1748	CCC-----	-----	1750
Dd	464	ProGlyLeu	ProLeuTryThrIleHisSerSerValAsnAspLysGlyLeuArgValLeu	483
OY	1751	---GACGACG	ACCCCTGCACAAAGCAAGCCCGCTTGGGCTAGCATGATGAGGACACC	1807
Dd	484	GlnAspAsn	SerAlaLeuAspLys-----MetLeuGlnAsnVal	496
OY	1808	AGCTGCCCCC	GGATTGTCTCTCCAGAAATCTTCATTTCACACGCGGCTCGGATGTG	1867
Dd	497	Gln-----	-----MetProSerLysLysLeuAspPheIleIleuAsnGlnIthr	511

```

0Y 1868 CGGGTC---TACGGCAGATCTACAAAGCCGCGGCTTCACGACGAGGAAAGACGCC 1924
Db 512 LysPheTrpTyrGlnMetIleLeuProPhe---PheAspLysSerLysLysTyrPro 530
0Y 1925 ACCGCTCTTTGATATATGAGAGCCGCCCAAGAGTCAGCTGGTAAATACCTCTCAAAAGC 1984
Db 531 LeuLeuLeuAspValTyrAlaGlyProCysSerGlnLysAlaAspThrValPhe----- 548
0Y 1985 ATCAAGTACTTGGCGGTCAAC-----ACACTGGCCCTCCCTGGGCTAGCGCGTGT 2035
Db 549 -----ArgLeuAsnTrpAlaThrTyrLeuAlaSerThrGlnAsnIleVal 564
0Y 2036 GTG---ATTACGGCAGAGGGCTCTGTCAGCAGAGGCTTCGGTTGAAAGGCGCTGAA 2092
Db 565 AlaSerPheAspArgTyrGlySerGlyTyrGlnLysLysIleMetHisAlaLeuAsn 584
0Y 2093 AACCAATATGGCCGCGTGAAGATCGAGACACAGGTGAGAGGGCTG---CAGTTCGTGGCC 2149
Db 585 ArgArgLeuGlyThrPheGluValGluAspGlnIleGluAlaAlaArgGlnPhe----- 602
0Y 2150 GAGAAGTATGGCTTCATCGACAGCTAGACCGAGTCCAGTCCATAGCTGGCTGTACAGGGGCG 2209
Db 603 SerLysMetGlyPheValAspAsnLysArgIleAlaIleTrpGlyTyrSerTyrGlyLys 622
0Y 2210 TTCCTCTGCGCATAGGGGGCTAAATCCACAAAGCCCAAGGTTCAGAGTGGCATCCGGGT 2268
Db 623 TyrValThrSerMetValIleGlySerGlySerGlyValPheLysCysGlyIleAlaVal 642
0Y 2270 GCCCGGCGTACCGCTGTAGTGGGCTGACGACACAGGTACACTAGACGAGGCTACATGACGTC 2329
Db 643 AlaProValSerAlaGlyTrpGlyTyrTyrAspSerValTyrThrGlnTyrGlyMetIleLeu 662
0Y 2330 -----CCTGGAACAACACAGACAGGCTATAGAGCGGAGTTCGCTGGCGCCCTGACGTGAG 2383
Db 663 ProThrProGluAspAsnLeuAspAsnLysTyrArgAsnSerThrValMetSerArgAlaGlu 682
0Y 2384 AAGTGGCCCAATGAGGCCAAACCGCTTGCTATCTCCACAGGCTTCCTGGACGAAGAGTGG 2443
Db 683 AsnPhe-----LysGlnValGluTyrLeuIleHisGlyThrAlaAspAsnVal 700
0Y 2444 CACTTTTCCACAAACTTCCTCGCTGCTCCCAACTGATCGAGGAGGAACCTTACAG 2503
Db 701 HisPheGlnGlnSerAlaGlnIleSerLysAlaLeuValAspValGlyAlaAspPheGln 720
0Y 2504 CTCGAGATCTACCCCAACAGAGACACAGATTCGCTGCCCGGAGTCGGGCGAGCATAT 2563
Db 721 AlaMetTrpTyrThrAspGluAspHisGlyIleAlaSerSerThrAlaHisGlnHisIle 740
0Y 2564 GAAGTCAACGTTGCTGCACCTTCTTACAGGA 2593
Db 741 TyrThrHisMetSerHisPheIleLysGln 750

RESULT 8
US-09-462-284-2
; Sequence 2, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Douglas, Agnes
; APPLICANT: Aftoller, Michael
; APPLICANT: Van Den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; TITLE OF INVENTION: ASPERGILLUS ORYZAE
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; CURRENT FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 771
; TYPE: PR1

```

```

; ORGANISM: Fungus
US-09-462-284-2

Alignment Scores:
Pred. No.:      2,89e-27      Length:      771
Score:          458.00      Matches:      173
Percent Similarity: 40.24%      Conservative: 95
Best Local Similarity: 25.98%      Mismatches: 268
Query Match:     9.55%      Indels:      130
DB:              4          Caps:      29

US-09-976-674-4 (1-2617) x US-09-462-284-2 (1-771)

QY 614 TCCTTCATCAATAACAGCAGCTGTGGGTGGCCAAATCAGACAGCGGAGCGGCGG 673
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 161 AlahValarValargluAsnAspLeuTyrVal-----TrpAspAsnGlyThrValThrArg 178

QY 674 CTGACCTTCCTGCCACCAAGGTTTATCCAAATGCTCTGGATGAC-----CCCAAGTCT 724
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 179 IleThr-----AspAspGlyGlyProAspMetPhe 188

QY 725 GCGGCTGGCGACCTTCGTATCATCAGAGAGATTG-----GACCGCTTCACGTGGGTAC 778
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 189 HisGlyValProAspTyrPileTyrGluGluIleLeuGlyAspArgTyr--AlaLeu 207

QY 779 TGGTGTGCCCCACAGCCTCTGTG-----GAAGGTTCAAGAGGCGCTCAG 823
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 208 TrpPheSerProAspGlyGluTyrLeuAlaTyrLeuSerPheAsnGluThrGlyValPro 227

QY 824 ACCGTGGGAATCCTGTATGAGAGATGATGATCCAGAGTGGAGGTGATTCACGTCCC 883
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 228 ThrTyrThrValGlnTyr---TyrMetAspAsnGluGluIle----- 240

QY 884 TCTCTCTCGCTAGAA---GAAGAGAGACGAGACTGTATCGGTACCCGAGAGAGCGAGC 940
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 241 AlaProAlaTyrProTyrPileGluLeuLys-----IleArgTyrProLysValSerGln 257

QY 941 AAGATCCCAAGATGCTGTGAAATGCGTGAAGTTCAGACTCAGACAGCAGCGGCAATC 1000
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 258 ThrAsnProThrValThrLeuSerLeu----- 266

QY 1001 GTCTCGACCCAGAGAGAGCTGTGTCACCCCTTCAGCTCGCTGTTCCGGAAGTGGAG 1060
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 267 LeuAsnIleAlaSerLysGluValLysGlnAlaProIleAspAlaPheGluSerThrAsp 286

QY 1061 TACATCGCCAGGCGCGGGTGGACCCGGGATGGCAATACCGCTGG----- 1105
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 287 LeuIleIle-----GlyGluValAlaTrpLeuThrAspThrHis 299

QY 1106 -----GCCATGTTCTGTGACCGGCCGCCAGCAGTGGCTCCAGCTCTCTCTC 1153
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 300 ThrThrValAlaAlaLysAlaPheAsnArgValGlnAspGlnGlnLysValAlaAlaVal 319

QY 1154 CCCCCGGCCCTGTTCATCCGAGACAGAGATGAGAGAGACGCGCTCAGCTCTGCCAGA 1213
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 320 -----AspThrAlaSerAsnLysAlaThrValIleSerAspArg 332

QY 1214 GCTGTCCCAAGATGTCACCGGTATGTGTGTACGAGAGAGTCAACCAAGCTGTGATC 1273
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 333 -----AspGlyThrAspGlyTyrPheLeu 339

QY 1274 ---AATGTCATGACATCTTCTATCCCTTCCCAATCAGAGAGAGAGAGAGCTCTGC 1330
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 340 AspAsnLeuLeuSerMetLysTyrIleGlyProIleLysProSerAspLysAspAlaTyr 359

QY 1331 TTTCTCCGCGCCATGATGATCAGACCGGCTTCGCCATTTGTGACAAAGTACCGCGCTT 1390
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 360 TyrIleAspLysSerAsp---HisSerGlyTrpAlaHisLeuTyrLeuPhe----- 375

QY 1391 TTTAAATCCAGGGCTACGATGTGAGTGAACCCCTTCAGCCCGGAGAGATGATTTAAG 1450
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 376 -----ProValSerGlyGly----- 380

```

```

QY 1451 TGCCCATTAAGAGAGATTTGCTGTGACACGCGGTGAATGGAGTTTGGCGAGGCAC 1510
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 381 -----GluProIleProLeuThrThrLysGlyAspTyrPileValThr----- 393

QY 1511 GGCTCCAAAGATCTGGGTGATGAGAGACCAAGCTGGTGTACTTCCAGGGCACCAGAGAC 1570
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 394 ---SerIleLeuSerIleAspGlnGluArgGlnLeuValTyrTyrLeuSerThrGlnHis 412

QY 1571 ACGCGCTGAGACACCACTTACGTGTGAGCTATGAG-----GGCGCGGCGAGATC 1624
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 413 HisSerThrGluArgHisLeuTyrSerValSerTyrSerThrPheAlaValThrProLeu 432

QY 1625 GATACGCTCACACGCGCGGCTTCCTCCATAGCTGTCCATGAGCCAGAACTTGACATG 1684
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 433 ValAspAspThrValAlaAlaTyr---TrpSerAlaSerPheSerAlaAsnSerGlyTyr 451

QY 1685 TTGCTCAGCCATTCACAGACCGGTGAGACCGCGCGCTGCGTGCACGTTTCAAAGCTGAC 1744
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 452 TyrIleLeuThrTyrGlyGlyProAspVal--ProTyrGlnGluLeuTyrThrThrAsn 470

QY 1745 GCGCCCGACGACGACCCCTGCACAAACGACCCCGCTTCGGGCTACATGATGGAGGCA 1804
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 471 SerThr-----LysProLeuArgThrIleThrAspAsnAlaLysVal 484

QY 1805 GCGAGCTGCCCGCGGATATGTTCTCCAGAGATCTTCATTTCAC-----ACGCGC 1858
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 485 LeuGluGlnIleLysAspTyrAlaLeuProAsnIleThrTyrPheGluLeuProLeuPro 504

QY 1859 TCGGATGTGGGCTCTACGGCATGATACAAAGCCCAAGCCTTGACCCAGAGAGAGAG 1918
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 505 SerGlyGluThrLeuAsnValMetGlnArgLeuProGlyPheSerProAspLysLys 524

QY 1919 CACCCCGCTGCTCTTGTATATGAGGCCCGCCAGGTGAGCGTGGATTAACCTCTTC 1978
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 525 TyrTrpIleLeuPheThrProTyrGlyGlyProGlyAlaGlnGluValThrLysArgTyr 544

QY 1979 AAGGCAATCAAGTAC---TTGCGGCTCAACACACTGCGCTCCCTGGGCTACGCGGTG 2035
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 545 GlnAlaLeuAsnPheLysAlaTyrValAlaIleSerAspSerGluLeuGluTyrValThrTrp 564

QY 2036 GTGATTCAGCGGACGGGCTCTGTGACCGAGGCTTCGTTCCGAGGGGCGCTGMAAAC 2095
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 565 ThrValAspAsnArgGlyThrGlnPheLysGlyArgLysPheArgSerAlaValThrArg 584

QY 2096 CAATATGGCCAGATGATGATCGAGACACGAGTGAAGGAGGCTGTGAGTTCGGCGCAGAG 2155
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 585 GlnLeuGlyLeuLeuGlnAlaGluAspGlnIleTyrAlaIleAlaIle---GlnAlaAlaAsn 603

QY 2156 TATGCTTCATGCACTGACCGCAGTTGCTCATGAGCTGTGCTCAAGGGGCTTCCTC 2215
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 604 IleProTyrPheAspAlaAspHisIleGlyIleTrpGlyTyrPheArgPheGlyGlyTyrLeu 623

QY 2216 TCGCTCATGGGCTATATCCCAAGCCCGCAGGTTCAGGTGGCCATCGGGGTGCCCG 2275
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 624 ThrSerLysValLeuGluLysAspSerGlyAlaPheThrLeuGlyAlaIleThrAlaPro 643

QY 2276 GTCACGCTGTGATGAGCTTCGACACAGAGGTGACACTGAGAGGCTACAGAGCTGCTGAG 2335
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 644 ValSerAspTyrPheArgPheTyrAspSerMetTyrThrGluArgTyrMetLysThrLeuSer 663

QY 2336 AACACCCAGCAGCGCTATGAGCGCGGTTCCGTCGCGCTGCACGTGAGAAAGTCCCAAT 2395
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 664 ThrAsnGluGluGlyTyrGluThrSerAlaValArg---LysThrAspGlyPheLysAsn 682

QY 2396 GAGCCCAACCGCTTGTATATCCACAGCGCTTCCTGAGCAAAACGTGACCTTTTCCAC 2455
    |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 683 ValGluGlyGlyPheLeuIleGlnHisGlyTyrGlyAspAspAsnValHisPheGlnAsn 702

QY 2456 ACAAACTTCTCGTCTCCCAACTGATCCGAGCAAAACCTTACCAAGCTCCAGATCTAC 2515
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 703 SerAlaAlaLeuValAspLeuLeuMetGlyAspGly-----ValSer 716

QY 2516 CCCAACGAGACACACAGT 2533

```

```

Db      717 ProGluLysLeuHisSer 722
      |||  :::  |||||
RESULT 9
US-08-230-491A-2
; Sequence 2, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Retlig, Wolfgang J.; Scanlan, Matthew J.;
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FELFE & LYNCH
; STREET: 805 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT - ASC II
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/230,491A
; FILING DATE: 20-APRIL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587299man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-230-491A-2

Alignment Scores:
Pred. No.:      8.64e-25      Length:      760
Score:          426.00      Matches:      168
Percent Similarity: 38.85%      Conservative: 109
Best Local Similarity: 23.56%      Mismatches:  256
Query Match:    8.88%      Indels:      180
DB:              1          Gaps:      31

US-09-976-674-4 (1-2617) x US-08-230-491A-2 (1-760)
QY      593 CCTCCGACCGCTGCTTCTTCTTCATCAATAAGACGACGTGTGGTGCCCAATC 652
      |||  |||||  |||
Db      176 ProGluLysProProPhe----- 181
QY      653 GAGACAGCGGAGGAGCGGCTGACCTTCGACCAAGGTTTATTCATGCTCGAT 712
      |||  :::::  |||  |||  |||
Db      182 -----GlnIleThrPhe-----AsnGlyArgGluAsnLysIlePhe 193
QY      713 GACCCCAAGTGTGGGGTGTGGCCACCTTCCTCATACGAGAAAGATTGACCGCTTCACT 772
      |||  :::::  |||  :::::  |||  :::::  |||  |||
Db      194 Asn-----GlyIleProAspTyrPalaTyrGluGluMet--LeuProThr 208
QY      773 GGGTAC-----TGTGTGTGCCCAACAGCCCTCTGGGAAGTTTCAGAGGCGCTCAAGACG 826
      |||  |||||  |||  :::::
Db      209 LysTyrIleLeuTyrPsrProAsnGlyLysPhe----- 220
QY      827 CTGGCAATCTGTATGAGGAAGTGCAGTGCAGGTGAGGTCAATTCAGTCCCTCT 886
      :::::  |||  |||  :::::  :::::  |||||
Db      221 -----LeuAlaTyrAlaGluPheAsnAspLysIleProValIleAlaTyrSerTyr 238

```

```

QY      887 CCTGCGCTAGAGAA-----AGGAAGACGACTCTATCGGTACCCAGACAGCAGCAG 943
      |||  :::::  |||  :::::  |||||  :::::  |||  |||
Db      239 TyrGluAspGluGluTyrProArgThrIleAsnIleProTyrProLysAlaGlyAlaLys 258
QY      944 AATCCCAAGATTGCTTGAACCTGCTGATGCCAGACTGACAGCAGCAGCAGCAAGATCGTC 1003
      |||||  :::::  |||  :::::
Db      259 AsnProValAlaArgIle-----PheIleIle 267
QY      1004 TCGACCCAGGAGAAAGAGCTGTGACGCCCTTCAGCTCGCTGTTCCGAGAGTG----- 1057
      |||  |||  |||  |||  |||
Db      268 AspThrThrTyrProAlaTyrValGlyProGluGluValProValProAlaMetIleAla 287
QY      1058 -----GAGTCATATCGCCAGCGGCGCGGTGAGACCGGATGGCAATATACCGTGGCC 1108
      |||  :::::  |||  |||  |||
Db      288 SerSerAspTyrTyrPheSerThrPheThrThrPalaThrAspGluArgValCys----- 305
QY      1109 ATGTTCTGGACCGCGCCACGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1168
      |||||  :::::  |||  |||||  :::::  |||
Db      306 -----LeuGlnThrPheLysArgValGlnAsnValSerValLeuSer 319
QY      1169 ATC-----CCGACACAGAGAT----- 1186
      |||
Db      320 IleCysAspPheArgGluAspTyrPglThrThrPaspCysProLysThrGluIleHisIle 339
QY      1187 GAGAGACAGCGGCTA-----GCCTCTGCCAGAGCTGCCAGAGATGTCACCGCTAT 1240
      |||||  |||  |||  |||  |||
Db      340 GluGluSerArgThrGlyTyrPalaGlyLysPhePheValSerAlaProValPheSerTyr 359
QY      1241 GTGTGTACGAGAGAGTCCACCAAGTGTGATCATGTATGACATCTTATCCCTTC 1300
      |||
Db      359 ----- 359
QY      1301 CCCCAATCAGAGGAGAGACGACGCTCTTCTCCGCCCAATGATGCAAGACCGGC 1360
      |||  :::::  |||  :::::  |||  |||
Db      360 -----AspAlaIleSerTyrTyrLysIlePheSerAspLysAspLys 373
QY      1361 TTCGCCATTGTACAAAGTCCAGCGCGCTTTAAATCCAGGGCTACGATGGAGTACG 1420
      :::::  |||||  :::::
Db      374 TyrLysHisIleHisTyrIle----- 380
QY      1421 CCTTCAGCCCGGGGAGATGATTAATTAAGTCCCATTAAGAAAGATGCTGTGAC 1480
      |||  :::::  |||  :::::  |||  |||
Db      381 -----LysAspThrValGluAsnAlaIleGlnIleThr 391
QY      1481 AGCGGTGAATGGAGGTTTGGCGAGCGACGCTCCAAAGATCTGGTCAATAGAGACGCC 1540
      |||||  :::::  |||  :::::
Db      392 SerGlyLysTyrPglAlaIle-----AsnIlePheArgValThrGluAsp 406
QY      1541 AAGCTGGTACTCTCCAGGACAC-----AAGACACAGCGCGCTGAGACACACCTCTAC 1594
      |||  |||  :::::  |||  :::::  |||
Db      407 SerLeuPheTyrSerSerAsnGluPheGluGluTyrProGlyArgArgAsnIleTyrArg 426
QY      1595 -----GTGTCAGCTATGAGCGCGCGAGATGAGTCAAGCTCTACACAGCCCGGCTTC 1648
      :::::  |||||  :::::  |||  |||
Db      427 IleSerIleGlySerTyrProProSerTyrLysCysVal----- 439
QY      1649 TCCCATAGCTCTCCATG-----AGCCAGAACTTGACATG 1684
      :::::  :::::
Db      440 -----ThrCysHisLeuArgLysGluArgCysGlnTyrThrAlaSerPheSerAsp 457
QY      1685 TTCGTACGCCACTACAGAGGCTGAGCAGCGCGCGCTCGTCCAGCTGTACACAGCTG--- 1741
      :::::  |||||  |||  |||  |||
Db      458 TyrAlaLysTyrTyrAlaLeuValCysTyrGlyProGlyIleProIleSerThrLeuIle 477
QY      1742 AGCGGCCCGGACGAGACCCCTGTCAC-----AAGCAGCCCGGCTTGTGGGTAGCATG 1795
      |||  |||  :::::
Db      478 AspGlyArgThrAspGluGlnIleLysIleLeuGluGluAsnLysGluLeuGluAsnAla 497
QY      1796 ATGAGCAGCAGCAGCTGCGCCCGGATTAATGCTCCAGAGATCTTCATTTCCACAGC 1855
      :::::  |||  |||  :::::
Db      498 LeuLysAsnIleGlnLeuProLysGlu-----GluIleLysLysLeuGluVal 513

```

```

QY 1856 CGCTGGATGTCGGCTTACGGCATGATCTACAGCCCGACGGCTTGACAGCGAG 1915
      ||| ||||| |||
Db 514 AspGluIlethrLeuTrpTyrLysMetIleLeuProboGln---PheAspArgSerLys 532
      ||| ||||| |||
QY 1916 AAGCACCACCGCTCCCTTTGATATGAGGCCCGACAGCTGCAGCTGGATTAATCTC 1975
      ||| ||||| |||
Db 533 LysTyrProLeuLeuIleGlnValTyrGlyGlyProCysSerGlnSerValArgSerVal 552
      ||| ||||| |||
QY 1976 TTC-----AAGGACATGACTCTTGGCGCTCAACACACTGGCTCCCTGGGCTACGCC 2029
      ||| ||||| |||
Db 553 PheAlaValAsnTrpIleSerTyrLeu-----AlaSerLysGluGlyMetVal 568
      ||| ||||| |||
QY 2030 GTGGTTGTGATGACGCGACGGGCTCCCTGACGAGGCGCTTGCTGGAAGGGCCCTG 2089
      ||| ||||| |||
Db 569 IleAlaLeuValAspGlyArgGlyThrAlaPheGlnGlyLysPryLysLeuTyrAlaVal 588
      ||| ||||| |||
QY 2090 AAAACCAATGGCGCCGAGATGAGATGAGACCGAGTGGAGGCCCTGCAGTTCGTGGC 2149
      ||| ||||| |||
Db 589 TyrArgLysLeuGlyValTyrGluValGluAspGlnIleThrAlaValArgLysPheIle 608
      ||| ||||| |||
QY 2150 GAGAAATGATGCTTATGACGCTGACGCGAGTTCAGTCCATGGCTGGTCTACGGGGCG 2209
      ||| ||||| |||
Db 609 Glu---MetIlyPheIleAspGlyLysArgIleAlaIleTrpGlyTyrPserTyrGlyGly 627
      ||| ||||| |||
QY 2210 TTCCTGCTGCTATGGGCTAATCCACAGCCCGAGTGTTCAGAGTGGCCATCCGGGT 2269
      ||| ||||| |||
Db 628 TyrValSerSerLeuAlaLeuAlaSerGlyThrGlyLeuPheLysCysGlyIleAlaVal 647
      ||| ||||| |||
QY 2270 GCGCCGCTGACGCTGATGGCTAGACAGAGGAGGAGGCTGAGCTGATGAGAGCTC 2329
      ||| ||||| |||
Db 648 AlaProValSerSerTrpGlyTyrTyrAlaSerValTyrThrGluAlaArgPheMetGlyLeu 667
      ||| ||||| |||
QY 2330 CCT-----GAGAACCAACGACGACGCGCTATGAGCGGGGTCCGTCGCTGACGCTGAG 2383
      ||| ||||| |||
Db 668 ProThrLysAspRasPheLeuGlnIleLysTyrLysAsnSerThrValMetAlaArgAlaGlu 687
      ||| ||||| |||
QY 2384 AAGTCCCAATGAGCCCAACCGCTTATGCTCCACGAGCTTCTGACAGAAACGTG 2443
      ||| ||||| |||
Db 688 TyrPheArgAsnValAsp-----TyrLeuLeuIleHisGlyThrAlaAspAspAsnVal 705
      ||| ||||| |||
QY 2444 CACTTTTCCACCAAACTTCCTGCTCCCACTGATGTCGAGCGAGGGAACCTTACAG 2503
      ||| ||||| |||
Db 706 HisPheIleAsnSerAlaGlnIleAlaLysAlaLeuValAsnAlaGlnValAspPheGln 725
      ||| ||||| |||
QY 2504 CTCGAGATCTACCCCAAGAGACAGACAGATTCGTCGCCCGGAGGCGC----- 2554
      ||| ||||| |||
Db 726 AlaMetTrpTyrSerArgGlnAsnHisGlyLeu-----SerGlyLeuSerThr 741
      ||| ||||| |||
QY 2555 GAGCACTATGAGTCACTGCTGCTGCTGCTTCTTACAGGAA 2593
      ||| ||||| |||
Db 742 AsnHisLeuTyrThrHisMetThrHisPheLeuGln 754
      ||| ||||| |||

```

```

; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5350.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-619-280A-2

Alignment Scores:
Pred. No.: 8,64e-25 Length: 760
Score: 426.00 Matches: 168
Percent Similarity: 38.85% Conservative: 109
Best Local Similarity: 23.56% Mismatches: 256
Query Match: 8.88% Indels: 180
DB: 1 Gaps: 31

US-09-976-674-4 (1-2617) x US-08-619-280A-2 (1-760)
QY 593 CTGCGCGACCTGCTCTTCTGCTTATCATATACAGCAGCAGCTGGGGTCCCAACATC 652
      ||| ||||| |||
Db 176 ProGlyAspProProPhe----- 181
      ||| ||||| |||
QY 653 GAGACAGCGCAGAGCGCGCGCTGACCTTGCCACCAAGTTTATCAATGTCCTGAT 712
      ||| ||||| |||
Db 182 -----GlnIleThrPhe-----AsnGlyArgGluAsnLysIlePhe 193
      ||| ||||| |||
QY 713 GACCCCAAGTCTGCGGCTGGCGCACCTGTCATACAGAGAGATTCACCCCTTCACT 772
      ||| ||||| |||
Db 194 Asn-----GlyIleProAspTrpValTyrGluGluGluMet---LeuProThr 208
      ||| ||||| |||
QY 773 GGGTAC-----TGCTGGTCCCGCCACAGCCTCTGGGAGGTTCAAGGGCTCAAGACG 826
      ||| ||||| |||
Db 209 LysTyrAlaLeuTrpTrpSerProAsnGlyLysPhe----- 220
      ||| ||||| |||
QY 827 CTGCGAATCTGATGAGAGATGATGATCCGAGGTGAGGTGATTCACGTCCCTCT 886
      ||| ||||| |||
Db 221 -----LeuAlaTyrAlaGluPheAsnAspLysAspIleProValIleAlaTyrSerTyr 238
      ||| ||||| |||
QY 887 CTGCGCTAAGAGAA---AGAGAGCGGACTCGATGCGATGCCAGGACAGAGCAGAG 943
      ||| ||||| |||
Db 239 TyrGlyAspGluGlnTyrProAlaGlyThrIleAsnIleProTyrProLysAlaGlyAlaLys 258
      ||| ||||| |||
QY 944 AATCCCAAGATTCCTTGAATAGCTGCTGAGTCCAGATCAGACAGCGCAAGATCGTC 1003
      ||| ||||| |||
Db 259 AspProValAlaArgIle-----PheIleIle 267
      ||| ||||| |||
QY 1004 TCGACCCAGAGAGAGAGCTGTCAGCCCTTCAAGTCGCTGTTCCCGAAGTG----- 1057
      ||| ||||| |||
Db 268 AspThrThrTyrProAlaTyrGlyProGlnGluValProValProAlaMetIleAla 287
      ||| ||||| |||
QY 1058 -----GAGTACATCGCGAGCGCGGGTGGACCGCGAGTGAATAGCTGCTGGGCC 1108
      ||| ||||| |||
Db 288 SerSerAspTyrTyrPheSerTrpLeuThrTrpValThrAspAlaArgValCys----- 305
      ||| ||||| |||
QY 1109 ATGCTCTGACCGCGCCAGCAGATGCTGCTGCTGCTCCCTCCCGCGGCTGTTG 1168
      ||| ||||| |||
Db 306 -----LeuGlnTrpLeuLysArgValGlnAsnValSerValLeuSer 319
      ||| ||||| |||
QY 1169 ATC-----CCGAGCACAGAGAAAT----- 1186
      ||| ||||| |||

```

```

Db      320  ILeCysAspPheArgGluAspTrpGlnThrTrpAspCysProLysThrGlnGlnHisIle 339
QY      1187  GAGAGAGAGAGGGCTA-----GCCTGCGCAGAGCTGTCCCGAGATGTCCACCGCTAT 1240
Db      340  GlnGluSerIleThrGlnIleGlyLysPhePheValSerAlaGlyProValPheSerThr 359
QY      1241  GTGGTACAGAGAGAGTACCAACGCTGTGATCAATGTTCATGATCTTATCCCTTC 1300
Db      359  -----
QY      1301  CCCCAATCAGAGGAGAGAGAGAGCTGTCTCTTCGCGCCCAATGATGACAGACCGC 1360
Db      360  -----AspAlaIleSerThrGlnIlePheSerAspLysAspGly 373
QY      1361  TTTCGCCATTGTTCACAAGTCACCGCGCTTTAAATCCAGGCTACGATGTGAGTGC 1420
Db      374  TyrLysHisIleHisTyrIle-----
QY      1421  CCTTCAGCCCGGGAAGATGATTAATTAAGTCCCATTAAGAGAGATGTCTGAC 1480
Db      381  -----LysAspThrValGlnAsnAlaIleGlnIleThr 391
QY      1481  AGCGTGATGGAGGTGGCGAGGAGCGCTCAAGATCTGGTCAATGAGAGAGAC 1540
Db      392  SerGlySerTrpGlnAlaIle-----AsnIlePheArgValThrGlnAsp 406
QY      1541  AAGTGGTGTACTTCCAGGGGACC---AAGACACGCGCGTGGAGACACACTTAC--- 1594
Db      407  SerLeuPheTyrSerSerAsnGlnPheGlnIleTyrProGlnArgAsnIleTyrArg 426
QY      1595  -GTGGTACAGTATGAGAGCGCGGAGATCGTACGCTCACCACGCGCGGCTTC 1648
Db      427  IleSerIleGlySerTyrProProSerLysCysVal-----
QY      1649  TCCCATGAGCTGCTCATG-----ACCCAGAACTTCGACATG 1684
Db      440  -ThrcysHisLeuArgLysGluArgCysGlnTyrThrAlaSerPheSerAsp 457
QY      1685  TTTCGTACGCCACTACAGAGCGTACAGACGCGCGCTGCGACGCTTCACACTG--- 1741
Db      458  TyrAlaLysTyrTyrAlaLeuValCysTyrGlyProGlnIleSerThrLeuHis 477
QY      1742  AGCGGCGCGGAGAGAGCGCGCTGCAC-----AAGCAGCGCGCTTCGTGGGTAGCATG 1795
Db      478  AspGlyArgThrAspGlnGlnIleLysIleLeuGlnGlnLysGlnLeuGlnAla 497
QY      1796  ATGGAGGACAGCAGCTGCGCGCGATTAATGCTTCACAGACTTCATTCACACAG 1855
Db      498  LeuLysAsnIleGlnLeuProLysGlu-----GlnIleLysLysLeuGlnVal 513
QY      1856  CGCTCGGATGTGGCGCTCTACGGCATATCTACAGCCCGCAGCTTCGACCGAGAG 1915
Db      514  AspGlnIleThrLeuTyrTyrLysMetIleLeuProGln---PheAspArgSerLys 532
QY      1916  AAGCAGCCCAACCGCTTCGTATATGAGAGCGCGCGAGCGTGCAGTGTGTAATAACTCC 1975
Db      533  LysTyrProLeuLeuIleGlnValTyrGlyProCysSerGlnSerValArgSerVal 552
QY      1976  TTC-----AAAGCATCAAGTACTCGCGTCAACACACTGCGCTCGGTACAGCC 2029
Db      553  PheAlaValAsnTrpIleSerTyrLeu-----AlaSerLysGlnGlnMetVal 568
QY      2030  GTGGTTGTATGACGCGAGGGCTCTGTACAGCGAGGCTTCGTTGGAAGGCGCTG 2089
Db      569  IleAlaLeuValAspGlyArgGlyThrAlaPheGlnGlnLysLeuTyrAlaVal 588
QY      2090  AAAAACAATGAGGCGAGGTGAGATGAGAGACAGGTGAGAGGCGCTGCATTCGTGCC 2149
Db      589  TyrArgLysLeuGlyValTyrGlnValGlnAspGlnIleThrAlaValArgLysPheIle 608
QY      2150  GAGAAGTATGGCTTCATCGACCTGAGCGAGTTGCCATTCATGCTGCTGCTACAGGCGC 2209

```

```

Db      609  Glu---MetGlyPheIleAspGluLysArgIleAlaIleThrGlnTyrSerTyrGlyLys 627
QY      2210  TTCCTCTCGCTCATAGGGGCTAATCCACAGCGCGAGTGTCAAGGTGGCCATGCGGGT 2269
Db      628  TyrValSerSerLeuAlaLeuAlaSerGlyThrGlyLeuPheLysCysGlyIleAlaVal 647
QY      2270  GCGCGGCTCACGCTGTGATGAGCTTACAGACAGGATGACACTGAGCGCTCATGAGCGTC 2329
Db      648  AlaProValSerSerTrpGlnTyrTyrAlaSerValIleThrGlnAlaPheMetGlyLeu 667
QY      2330  CTT-----GAGAACAAACAGACAGCGCTATGAGCGGGGTTCCTGCGCTCAGCGTGCAG 2383
Db      668  ProThrLysAspAspAsnLeuGlnIleTyrLysAsnSerThrValMetAlaArgAlaGln 687
QY      2384  AAGTGGCCAAATGAGCCCAACCGCTGTCTATCTCCACGCGCTTCGTGAGCAAAAGCTG 2443
Db      688  TyrPheArgAsnValAsp-----TyrLeuLeuIleHisGlyThrAlaAspAspAsnVal 705
QY      2444  CACTTTTCCACACAACTTCCTGCTCCCACTGATCCGAGCAGGAGAAACCTTACAG 2503
Db      706  HisPheGlnAsnSerAlaGlnIleAlaLysAlaLeuValAsnAlaGlnValAspPheGln 725
QY      2504  CTCAGATTAACCCCAACAGAGACAGACAGTATTCGCTGCGCGCGAGTGGCGC----- 2554
Db      726  AlaMetTrpTyrSerAspGlnAsnHisGlyLeu-----SerGlyLeuSerThr 741
QY      2555  GAGCAGTATGAGTACAGCTGCTGCTCATCTTCTACAGGAA 2593
Db      742  AsnHisLeuTyrThrHisMetThrHisPheLeuLysGln 754

RESULT 11
US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.;
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

US-08-940-391-2

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:
Score:	8,64e-25	760	168	109	256
Percent Similarity:	426.00				
Best Local Similarity:	38.85%				
Query Match:	23.56%				
DB:	8.88%				
	2	Gaps:	31		

US-09-976-674-4 (1-2617) x US-08-940-391-2 (1-760)

```

QY 593 CCGGCGACACCTGCTTCTCTCTCATATACAGCAGCCTGTGGTGGCCACATC 652
    ||| ||||| |||
Db 176 ProGlyAspProPhe-----
QY 653 GAGACAGCAGGAGGCGGCTGACCTTCGCCACAGGTTTATCCAAATGCTGGAT 712
    ||| ||||| |||
Db 182 -----GlnIleThrPhe-----AsnGlyArgGluAsnLysIlePhe 193
QY 713 GACCCCAAGTCTGGGGGTGGCCACCTTCATACAGAGAGAGTTCAGCCGCTTCACT 772
    ||| ||||| |||
Db 194 Asn-----GlyIleProAspTrpValTyrGluGluMet-----LeuProThr 208
QY 773 GGGTAC-----TGGTGGTCCCCACAGCCTCTGGGAGGTTTCAGAGCCTTCAGAGC 826
    ||| ||||| |||
Db 209 LysTrpIleLeuTrpTrpSerProAsnGlyLysPhe-----
QY 827 CTGGAACTCTGATGAGAGTGCATGATCGAGGTGGAGGTCATTCAGTCCCTCT 886
    ||| ||||| |||
Db 221 -----LeuAlaTrpAlaGluPheAsnAspLysAspIleProValIleAlaTyrSerTyr 238
QY 887 CCGTGGCTAGAGAA--AGGAAGACGAGTCTGATGCCACCCAGAGCAGCAGCAG 943
    ||| ||||| |||
Db 239 TyrGlyAspGluGlnTyrProArgThrIleAsnIleProTyrProLysAlaGlyAlaLys 258
QY 944 AATCCCAAGATTGCTTGAATAGTGGTGAAGTCCAGACTACAGTACAGCCAGGCAAGATGTC 1003
    ||| ||||| |||
Db 259 AsnProValAlaArgIle-----PheIleIle 267
QY 1004 TCGACCCAGAGAGAGAGTGTGTCAGCCTTCAGTCCGCTGTTCCGAGAGTG- 1057
    ||| ||||| |||
Db 268 AspThrThrTyrProAlaTyrValGlyProGlnGluValProValProAlaMetIleAla 287
QY 1058 -----GAGTACATCGCCAGGCGCGGTGGAACCGGAGTGCAGCAATCGCCTGGCC 1108
    ||| ||||| |||
Db 288 SerSerAspTyrTyrPheSerTrpLeuThrTrpValThrAspGluArgValCys----- 305
QY 1109 ATGTCTGTGACCGGCCCCCGACAGTGGCTCAGCTGCTCTCCCTCCCGGCCCTGTC 1168
    ||| ||||| |||
Db 306 -----LeuGlnTrpLeuLysArgValGlnAsnValSerValIleuSer 319
QY 1169 ATC-----CCGACACAGAGAAAT----- 1186
    ||| ||||| |||
Db 320 IleCysAspPheArgGluAspTrpGlnThrTrpAspCysTrpLysThrGlnGlnHisIle 339
QY 1187 GAGGACAGCGGCTA-----GCCCTGCCAGAGCTGTCCCGAGGAATGTCAGCCGCTAT 1240
    ||| ||||| |||
Db 340 GluGlnSerArgThrGlyTrpAlaGlyGlyPhePheValSerArgProValPheSerTyr 359
QY 1241 GTGTGTAGAGAGAGGTCACCAACGTCGTGATCAATGTTATGACATCTTCTATCCCTTC 1300
    ||| ||||| |||
Db 359 ----- 359
QY 1301 CCCCATAACAGGAGAGAGAGAGCTGCTTCTCCGCGCAATGATGACAGACCGGC 1360
    ||| ||||| |||
Db 360 -----AspAlaIleSerTyrTyrLysIlePheSerAspLysAspGly 373
QY 1361 TTCTGCATTGTGACAAAGTACCGCGCTTTAAATCCAGGAGCTACGATTGGAGTGAG 1420
    ||| ||||| |||
Db 374 TyrLysHisIleHisTyrIle----- 380
QY 1421 CCCTTCAGCCCCGGGAGAGATGAATTTAAGTCCCATTTAAGAGAGAGATTGCTTGACC 1480

```

```

Db 381 -----LysAspThrValGluAsnAlaIleGlnIleThr 391
QY 1481 AGCGTGATGGAGAGGTTTGGCGAGCAGCGGCTCCAAATCTGGGTCATGAGGAGACC 1540
    ||| ||||| |||
Db 392 SerGlyLysTrpAlaIle-----AsnIlePheArgValThrGlnAsp 406
QY 1541 AACGTGGTACTTCCAGGCGACC--AGGACAGCGCGCTGGAGACACACCTTAC-- 1594
    ||| ||||| |||
Db 407 SerLeuPheTyrSerSerAsnGlnPheGluGluTyrProGlyArgArgAsnIleTyrArg 426
QY 1595 -----GTGGTACGCTATGAGCGCGCGGAGATGCTTACGCTCAGCAGCGCGGCTTC 1648
    ||| ||||| |||
Db 427 IleSerIleGlySerTyrProSerLysLysCysVal----- 439
QY 1649 TCCCATGACTGTCTTCACG-----AGCGAAGCTTCGACATG 1684
    ||| ||||| |||
Db 440 -----ThrCysHisLeuAlaGlyLysGluAlaGlyGlnIleTyrThrIleAlaSerPheSerAsp 457
QY 1685 TTGTCAGCCCATACAGACAGGTGAGACAGCGCGCGCTGCGACGCTTCAAGCTG-- 1741
    ||| ||||| |||
Db 458 TyrAlaLysTyrGlyAlaLeuValCysTyrGlyProGlyIleProIleSerThrIleHis 477
QY 1742 AGCGCGCCGACAGACAGCCCTGAC- -AAGCAGCGCGCTTCTGGCTAGCATG 1795
    ||| ||||| |||
Db 478 AspGlyArgThrAspGlnGluIleLysIleLeuGlnLysAsnGlyLysGluAsnAla 497
QY 1796 ATGAGGACAGCAGCAGTGGCCCCCGGATTAATGTTCTCCAGAGATCTTCATTTCCACAGC 1855
    ||| ||||| |||
Db 498 LeuLysAsnIleGlnLeuProLysGlu-----GlnIleLysLysLeuGluVal 513
QY 1856 CGCTCGATGTGGCGCTCTACGCGCATGATACAGCCCGCAGCCTTCAGCAGGAGAG 1915
    ||| ||||| |||
Db 514 AspGlnIleThrLeuTrpTyrLysMetIleLeuProGln--PheAspArgSerLys 532
QY 1916 AAGCAGCCCGCAGCTCTTGTATATGAGCGCCCGCAGCTGCACTGGATTAATCTCC 1975
    ||| ||||| |||
Db 533 LysTyrProLeuLeuIleGlnValTyrGlyLysProCysSerIleSerValArgSerVal 552
QY 1976 TTC-----AAAGCATCAAGTACTTGGCGCTCAACACACTGGCTCCCGGCTACGCGC 2029
    ||| ||||| |||
Db 553 PheAlaValAsnTrpIleSerTyrLeu-----AlaSerLysGluGlyMetVal 568
QY 2030 GTGGTGTGATGACGCGCAGGCGCTCTGTACAGCAGGCGCTTGGTTCAGAGCGCCCTG 2089
    ||| ||||| |||
Db 569 IleAlaLeuValAspGlyArgGlyThrAlaPheGlnGlyAspLysLeuLeuTyrAlaVal 588
QY 2090 AAAAACCAATGGCGCAGGTGAGATCGAGACACAGTGGAGGCGCTGCAAGTTCGTGGCC 2149
    ||| ||||| |||
Db 589 TyrArgLysLeuGlyValTyrGluValGluAspGlnIleThrAlaValArgLysPheIle 608
QY 2150 GAGAAGATGAGCTTTCATGACGTCGAGCGGAGTGCATGCACTGAGTGGTCTAGCGGCGC 2209
    ||| ||||| |||
Db 609 Glu--MetGlyPheIleAspGluLysArgGlyIleAlaIleTrpGlyTrpSerTyrGlyGly 627
QY 2210 TTCTCTGCTCATGGGGCTAATTCACAAAGCCCAAGGTTCACAGGCTTCAAGGCTTCCGCGGT 2269
    ||| ||||| |||
Db 628 TyrValSerSerLeuAlaLeuAlaSerGlyThrGlyLeuPheLysCysGlyIleAlaVal 647
QY 2270 GCCCGGTCACGCTGAGGAGGCTTACGACAGAGGATGACATGAGGCTTCAAGAGAGCTC 2329
    ||| ||||| |||
Db 648 AlaProValSerSerTrpGlyTyrAlaSerValTyrThrIleTrpPheMetGlyLeu 667
QY 2330 CCT-----GAGAACAACAGCAGCAGGCTATGAGCGGGTTCCTGCGCCCTGACGTTGAG 2383
    ||| ||||| |||
Db 668 ProThrLysAspAsnLeuGlnHisTyrLysAsnSerThrValMetAlaArgAlaGlu 687
QY 2384 AAGTGGCCCAATGAGCCCAACGCTTGCCTTATCTCCAGCGCTTCTGACGAAAACTG 2443
    ||| ||||| |||
Db 688 TyrPheArgAsnValAsp-----TyrLeuLeuIleHisIleThrAlaAspAsnVal 705
QY 2444 CACTTTTCCACAAACCTTCTCGTCTCCCAACTGATCGAGCAGGAAACCTTACAG 2503
    ||| ||||| |||

```


Db 706 HispheglinaSerAlaGlnIleAlaLysAlaLeuValaAsnAlaGlnValaAspHeGln 725
QY 2504 CTCGAGATCTACCCCAAGAGACAGATATTCGCTCCCGAGCTGGGC----- 2554
Db 726 AlaMetIprYrSerAspGlnAsnHisGlyLeu-----SerGlyLeuSerThr 741
QY 2555 GAGCAGTATGAAGTCACGTTGCTGCACCTTTCTACAGGAA 2593
Db 742 AsnHisLeuYrThrHisMetThrHisPheLeuLysGln 754
RESULT 12
US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent NO. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOORROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHIELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-5
Alignment Scores:
Pred. No.: 8,39e-12 Length: 960
Score: 259.00 Matches: 232
Percent Similarity: 28.90% Conservative: 22
Best Local Similarity: 26.39% Mismatches: 340
Query Match: 5.37% Indels: 285
Gaps: 50
US-09-976-674-4 (1-2617) x US-09-219-849-5 (1-960)
QY 2432 CCAGAGACCTCGAGAGTAAGCAAGCGTTGGGCTCATTTGGCGACGCTTCCACGATCA 2373
Db 61 ProGlySerArgAspProGlyProGlyProAlaGlyProAlaGlyProProGlySer 80
QY 2372 -----GGGCCAGGAAACCGCGCTCATGAGCGCTGTGT 2340
Db 81 ArgAspProGlyProProGlyAlaAlaProGlyProAlaGlyProProGlySerArgAspPro 100
QY 2339 TGTTCAGGAGCATGTCATGTAGCGCTCAGTACCTGTGTGTGAGGCCATCCAGACGG 2280
Db 101 GlyProGlyAlaAlaProGlyProAla-----GlyProProGlySerArg 115
QY 2279 TGACCGGGGACACCGCGATGGCCACTTGAACACCTGGGGCTGTGGATTAGCCCCATGA 2220
Db 116 AspProGly-----ProProGlyAlaAlaProGlyProAlaGly---ProProGly 130
QY 2219 GCGAGA-----GGAGCGCCCGTAGACACACCATGATGAGCACTCGGCTCAGATGCA 2166
Db 131 SerArgAspProGlyProProGlyAlaAlaProGlyProAlaGlyProProGlySerArgAsp 150
QY 2165 TGAAGCCTACTCTCGGCCACGAAGTCAAGCCCTCA----- 2127
Db 151 ProGlyProProGlyAlaAlaProGlyProAlaGlyProProGlySerArgAspProGlyPro 170

QY 2126 -----CGTGGCTCGATCTCA-----CGTGGCCATTGGCTTT 2091
Db 171 ProGlyAlaAlaProGlyProAlaGlyProProGlySerArgAspProGlyProGlyAla 190
QY 2090 TCAGGGCCCTTCGAAACGAAGCCCTCGTAGACAGACCCCTGCGGCAATCAACACCA 2031
Db 191 ProGlyProAlaGlyProProGlySerArgAspProGlyProProGlyAlaAlaProGlyPro 210
QY 2030 -----CGCGTAAGCCAGGAGGCGCATGTGTGAGCCGCAACTTGA 1986
Db 211 AlaGlyProProGlySerArgAspProGlyProPro----- 222
QY 1985 TGCCTTGAAGAGTATTCACAGCTGCACCTACCGGCGCTCCATATCAAAAGAGACGG 1926
Db 223 -----GlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAlaGlyPro 239
QY 1925 TGGGTGTCTCTCCCTCGTCGCAAGGCGTGAGGCTTGTAGATCATGCCGTAGACCCCA 1866
Db 240 LysGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAlaGlyProLysGly 259
QY 1865 CATCCGAGCGCGTGTGAAGATGGAAGATCTGTGAGGAACATATCCGGG-----GCG 1812
Db 260 AlaProGlyProAlaGlyProProGlySerArgAspProGlyProProGlyAlaAlaProGly 279
QY 1811 AGCTGGCTGCTCCATCTAGTCTAGCCCAAGAGCGGCGCTGCTTGCAAGGGGCTGTCT 1752
Db 280 ProAlaGlyProPro-----GlySerArgAsp 288
QY 1751 CGGGCCGCTCAGCTTGTAGACGTGCACGACAGGCGGCGGTGCTCAGCTGTATGTGCG 1692
Db 289 ProGly----- 290
QY 1691 TGACGACATGTCGAAGTTTGGCTCATGAGACACTATGGAGAGACCGGCGGTGTGA 1632
Db 291 -----ProProGlyAlaAlaProGlyProAlaGlyProPro----- 301
QY 1631 GCGCAGCATCTCGCCGCGCGCTCATAGCTGACACAGCTAGAGGTGTGTCCAGCGCG 1572
Db 302 GlySerArgAspProGlyProProGlyAla---ProGlyProAlaGlyProProGlySer 320
QY 1571 -----TGTCCTGGCGCTGGAAGTACACAGCGTGTGTCTCATATGACCC 1524
Db 321 ArgAspProGlyProProGlyAlaAlaProGly-----ProAlaGlyProPro----- 335
QY 1523 AGATCTTGAGACCGCTGCTCGCAAAACCTCCATTCACCGCTGTGACAGCAATCTCTT 1464
Db 336 -----GlySerArgAspProGlyProPro----- 343
QY 1463 CCTTAATGGGCACTTAATTCATCTTCCCGGGGCTAAGGGCTCACTCCAAATCGTAGC 1404
Db 344 ----- 345
QY 1403 CCTGGATTTTAAACGCGGAGACTTGTACAATGCGAGAGCGGCTTGCATTGAT 1344
Db 346 ProGly----- 347
QY 1343 TGCGCGGAGAAAGCAGAGCTGCTCTCTCCCTGATTTGGGGAGAGATAGAAGATGT 1284
Db 348 -----ProAlaGlyProProGlySerArgAspProGly--- 358
QY 1283 CATGAACATTATCCAGACGTTGGTGTGACCTCTCTGTACACACATAGCGCTGAGACATTCC 1224
Db 359 -----ProProGlyAlaAlaProGlyProAlaGlyProPro 369
QY 1223 TGGGACAGCTCTGACAGAGCTAGCCGCTGCTCTATTCCTGTGTGCGGATGAGACA 1164
Db 370 ---GlySerArgAspProGlyProProGlyAlaPro-----GlyProAla 383
QY 1163 GGGCCGGGGGAGGA-----GACAGAGCTGAGAGCCACTGTGGGCGCGGTCCAGGA 1113
Db 384 GlyProProGlySerArgAspProGlyProProGlyAlaAlaProGlyProAlaGlyPro--- 402

```

QY 1112 ACATGCCAGGCGATTATTCGCAATCCGGGTGCACCGGCGCATGTACTCCACCT 1053
    |||||
Db 403 -----ProGlySerArgAspPro-----GlyProPro 411
QY 1052 TCGGGAACAGCGACTAGAGGGCTGCACAGCTCTTCTCTGGGTGCAGACGATTTTGC 993
    |||||
Db 412 G1yAla-----ProGlyProAlaGlyProProGlySerArgAspProGly 426
    |||||
QY 992 CCTGGCTGTGAGTGTGAACT-----CAGCCAGTTTCAAGCAATCTTGGGATTTCTGC 939
    |||
Db 427 ProProGlyAlaProGlyProAlaGlyProProGlySerArgAspPro----- 442
QY 938 TGCCTGTCTGGGTACCAGATACAGTCCGCTTCTTCTTACGCGCAGAGAGGGGA 879
    |||
Db 443 -----GlyProProGlyAlaProGlyPro-----AlaGlyProProGlySer 456
    |||
QY 878 CGTGAATGACCTCCACTCCGAGCTCAT-----CGACTTCCCAT----- 840
    |||
Db 457 ArgAspProGlyProProGlyAlaAlaHisGlyProAlaGlyProGlyAlaHisGlyPro 476
    |||
QY 839 ACAGGATTCGCGAGCGTGTGAGGCTTGCAGACCTTCCAGAGAGGCTGTGGGCGACAC 780
    |||
Db 477 AlaGlyProGlySerGlyAlaHisGlyProAlaGly-----ProGlySerGlyAlaHisGlyProAla 495
    |||
QY 779 AGTACCAGTGAACG---GTCGAACTTCTCTGTATGACGAGGTGGCCACACCCGAG 723
    |||
Db 496 G1yProGlySerGlyAlaProGlyProAlaGlyProProGlySerArgAspProGlyProPro 515
    |||
QY 722 ACTTGGGGTCTATCCAGACATGTGATAACCTTGTGGCAGAAAGTCAAGCCCGCGCTCT 663
    |||
Db 516 -----GlyAlaProGly-----ProAlaGlyPro 523
    |||
QY 662 CGCCTGTCTGCAATTTGGGCCACCCACAGTGCCTGTATTGATGAGAGAGAAGAGCAG 603
    |||||
Db 524 ProGlySerArgAspProGlyProProGly-----AlaPro 535
QY 602 GGTCCGAGAGGC-----AGATTTGGGCTCCAGCCGCGCCCTGAGCAGCTGGG 555
    |||
Db 536 G1yProAlaGlyProProGlySerArgAspProGlyProProGlyAlaProGly----- 553
    |||
QY 554 TCTTGATTTCCAGGGGTTTCTATAGGGGACACCATGAAGCCCTTCTGGCGCGCTCCGAG 495
    |||||
Db 554 -----ProAla-----GlyProProGlySer-----ArgAsp 562
    |||
QY 494 AGTGAAGAGGCTGTGCTGCTGAGTGAAGAGAGAGCCACTCTCGCTGTGGAAGTCTGT 435
    |||
Db 563 ProGlyPro-----ProGlyAlaProGlyProAlaGlyProProGlySerArg 578
    |||||
QY 434 AGGAGGTGATGCCGAGACACCCAGCGGCTTCCGCTCCAGCAGCTCTCTCCGAG 375
    |||||
Db 579 Asp-----ProGlyProProGly-----AlaProGlyProAlaGlyProProGly 593
    |||||
QY 374 AGTAGACCCCATGTGGGGGCTGGCTGGAATATCATCATGTGCTTCCAGAGACAGA 315
    |||||
Db 594 SerArgAspProGly-----ProProGlyAlaProGlyProAlaGlyProProGly 610
    |||
QY 314 GCACAGAGGCT-----CTTTCGACCTTCTTGGGAATCTCAAGATGAGAGA 267
    |||
Db 611 SerArgAspProGlyProProGlyAlaProGlyProAlaGlyProProGlySerArgAsp 630
    |||
QY 266 GGGAGTTCTCTGGCTGC-----CATATGGCATTTCCAGTAGTAAGGC----- 222
    |||
Db 631 ProGlyProProGlyAlaProGlyProAlaGlyProProGlySerArgAspProGlyPro 650
    |||
QY 221 ---GCTGGAGTGGGGCCAGACTCATCCGCTTCTGCACAAATGGAAGTCGGGGCG 165
    |||
Db 651 ProGlyAlaProGlyProAlaGlyProProGlySerArgAspProGlyProProGlyAla 670
    |||
QY 164 CTTGTGTGACAATGAGCGCGAGT---ACTTGGCGGCTGCGGTGATGATGCTCGGAGCC 108
    |||
Db 671 Pro-----GlyProAlaGlyProProGlySerArgAspProGlyProProGly 686
    |||
QY 107 CGTCCACGAGTGTCTTGCACTGGAAGCGGGCGCGGCTCATCTGTGC----- 56

```

```

Db 687 AlaProGlyProAlaGlyProProGlySerArgAspProGlyPro-ProGlyAlaHisGly 706
    |||
QY 55 -----GGCTGGCTCCGCTCCGTCGCGCGTGGGGTCCCGGTGTCGCATGCT 8
    |||
Db 706 yProAlaGlyProGlySerGlyAlaHisGlyProAlaGlyProGlySerGlyAlaHisGly 724
    |||
RESULT 13
US-08-642-255-62
; Sequence 62, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CARPELLO, Joseph
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1064 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-62
Alignment Scores:
Pred. No.: 2,57e-11 Length: 1064
Score: 253.00 Matches: 228
Percent Similarity: 27.64% Conservative: 18
Best Local Similarity: 25.62% Mismatches: 330
Query Match: 5.25% Indels: 315
DB: 1 Gaps: 48
US-09-976-674-4 (1-2617) x US-08-642-255-62 (1-1064)
QY 2516 GGTAGATCTGAGACT---GGTAGGTTTCCCTCGGATCATGTGGAGACGAGAGT 2460
    |||
Db 103 G1yProProGlyAlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyPro 122
    |||
QY 2459 TTGCTGGAAGAAAGTGCACGTTTTCGTCACGGAAGCCGTGGAGATACGACGCGTTGG 2400
    |||
Db 123 ProGlyAlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGly 142
    |||
QY 2399 GCTCATTTGGCAGCGCTTCTCCA-----CGTGCA 2373
    |||
Db 143 AlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAlaPro 162
    |||
QY 2372 GGGCCAGGAACCGCTCATATAGCCGTGTGTTGTTCTCAGGAGCGTCCATGTAGCGCT 2313

```


QY 255 CGGCTGCATATGCAATCCAGCTAGAGGC-----GGTGGAGTGGGGCCAGAC 202
Db 647 aProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAla----- 665
QY 201 TCATCCGCTCTTGCAC---CAAACTGGAAGTCGTGGGGCCCTGTTGACATATAGGCC 145
Db 666 ---ProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAlaPr 684
QY 144 GAGTACTGGGGCTGGCGGTGATGATGTCGCGAGCCGCCGACGAGTGGCTTCTGGACC 85
Db 684 OGlyProAlaGlyProProGly-----SerArgGlyAspProGlyProProGlyAlaPr 702
QY 84 TGGAC---AGCGGGGGCCGGGTATCTGTGAC-----GGCTGGCGCTCGGTGCGCC 34
Db 702 OGlyProAlaGlyProProGlySerArg-GlyAspProGlyProProGlyAlaProGlyP 722
QY 33 GTTGGGCTCCGGTGGTGGCCATGT 8
Db 722 roAlaGlyProProGlySerArgGly 730

RESULT 14
US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chiviva, John
; APPLICANT: Vaccik, Peter
; TITLE OF INVENTION: SNE2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1998-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRP
; ORGANISM: Human
US-09-579-181-2

Alignment Scores:
Pred. No.: 7,94e-11 Length: 2972
Score: 249.50 Matches: 242
Percent Similarity: 35.64% Conservative: 87
Best Local Similarity: 26.22% Mismatches: 371
Query Match: 5.20% Indels: 224
Db: 4 Gaps: 47

US-09-976-674-4 (1-2617) x US-09-579-181-2 (1-2972)

QY 67 CCGGGCCGGCGGCTTCAGAGTCGAGAGCACTGCTGGAGGGCTCGGAGCATCATCA 126
Db 826 ProValArgProProGlyProGlu-----LeuSerAlaGlnProThrProGlyPro 843
QY 127 CGGACGCGGCAAGTACTCGGGCCCTCATATTGCAACAAGCGCCGCCACGACTTCCAGTTGT 186
Db 844 ---ValProGlnValLeuProAlaSerLeuMetValSerLa----- 856
QY 187 GCAGAGACGAGATGACTGTGGCCGCCACCTCCACCGCGCTTACTACTGGGAATGCCATA 246
Db 857 -----SerProAlaGlyProProGlyLeuLeuProAlaSerArgPro 869
QY 247 TGGAGCGGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAAAGTCCGGAAGAAGGC 306
Db 870 ProGlyProValLeuLeuProProLeu-----GlnProAsnSerGly 883
QY 307 TCTGTGCTCTCTCTCTCGAAGAGATGATCATTTCCAGGCCAGCC---CCACCA 363
Db 884 SerLeuProGlnValLeuProSerProLeuGlyValLeuSerGlyThrSerArgProPro 903
QY 364 TGGGGTCTACTCTCG-----GGAGGAGGAGCTGCTGAGGAGCGGAA 405

Db 904 ThrProThrLeuSerLeuLysProThrProProAlaProValArgLeuSerProAlaPro 923
QY 406 ACCGCTGGGGG-TCTTGGGATACCTCTACGACCTTCCACAGGAGAGGGCCTCTCC 464
Db 924 ProProGlySerSerSerLeuLeuLysProLeuThrValProProGlyTyrThrPhePro 943
QY 465 TCTTCCAGGCGCAGACAGCCCTTTCCTCTGCTCGCGAGCGCGGACAGAGGCGCTTCATGG 524
Db 944 ProAlaAlaAlaThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 961
QY 525 TGTCCCTTAAGAACCCTGGAATCAAGAACCAAGTCTCAAGGGCCCGGATGAGACCCA 584
Db 961 aValProAlaProThrProAlaProGlnArgLeuLeuLeu-----SerProAspMetGlnAl 980
QY 585 AAATGTCCCTCCGCGACCGCCCTT-----CTTCTCTCTTCATCAATAACAGCGACC 635
Db 980 aArgLeuProSerGlyGlyValValSerIleGlyGlnLeuAlaSerLeuAlaGlnArgPro 1000
QY 636 TGTGGGTGGCCACATCAGACAGGAGGAGCGCGCGGTGACCTTCTGCCACCAAGTT 695
Db 1000 oval-Ala-----AsnAlaGlyGlySerLysProLeuThrPheGlnIleGlnIleVal 1017
QY 696 TATCCAATGTCTG-----GATGACCCCAAGTCTG 725
Db 1017 snLysLeuThrLeuThrGlyAlaGlnValArgGlnLeuAlaValGlyGlnProArgProL 1037
QY 726 CGGCTGGCGCCACTTCTGCTATACAGAAAGTTCGACCGCTTACAGGCTACGTGGGT 785
Db 1037 euGlnMet-ProProThrMetValAsnAsnThrGlyValIleValIleValIleValArgGln 1056
QY 786 GCCCAGCAG-----CCTCTGGGAAGTTCAGAGGCGCTCAAGACGCTGGCAA 833
Db 1057 AlaProArgAspGlyLeuThrProValProProLeuAlaProAlaArgProProSer 1076
QY 834 TCTGTATGAGGAAGTGTGATGATCGAGGTGAGGTGATTCACGTCCCTCTCTGCGC 893
Db 1077 SerGlyLeuProAlaValLeuAsnProArg-----ProThrLeu--- 1089
QY 894 TAGAAGAAAGGAAGAGGAGCTGATGCGTATCCCGACAGACAGGAGCAAGATCCCAAGA 953
Db 1090 -----ThrProGly-----Arg 1093
QY 954 TTGCTTGAACATGCGTACGTTCACAGTACAGACCGGCGCAAGATGCTT----- 1004
Db 1094 LeuProThrProThrLeuGly-----ThrAlaArgAlaProMetProThrProThr 1110
QY 1005 -----CGACCCAGAGAGAGAGCTGTGACGCCCTTACGCTGCTTCCGAGGTGG 1058
Db 1111 LeuValArgProLeuLeuLysLeuValHisSerProSerProGlyValSer----- 1127
QY 1059 AGTACATGCGCAGGCGCGGGTGGACCGCGGATGGCAATACGCGTGGCATGTCTCTGG 1118
Db 1128 -----AlaSerAlaProGlyAlaAlaProLeu 1136
QY 1119 ACCGCGCCAGCAGATGCTCAGCTGCTCTCTCCCGCGCGCTTATCCCGAGCA 1178
Db 1137 ThrIleSerSerProLeuHisValProSerSerLeuProGlyProAlaSerSerProMet 1156
QY 1179 CAGAGATGAGAGACAGCGGCTAGCCTTCTGCGAGAGCTGTCCCGAGAAATGTCAGCGGT 1238
Db 1157 -----ProIleProAsnSerSerPro-----LeuAlaSerPro 1167
QY 1239 ATGAGTGTACGAGAGAGTCCACCAAGCTGTGATCAATG-----TTCATGACATCTCT 1292
Db 1168 ValSerSerThrValSerValProLeuSerSerSerLeuProIleSerValProThrThr 1187
QY 1293 ATCCCTTCCCCCAATCAGAGGAGAGCAGCTGCTTCTTCCGCGCAATG----- 1346
Db 1188 LeuProAlaProAlaSerAlaProLeuThr-----IleProIleSerAlaProLeuThrVal 1206
QY 1347 AATGCAAGACCGGCTTCTGCTGCAATTTGTACAAAGTACACCGCGTTTAAATCCAGGGCT 1406
:::|||||
:::|::|

```

Db 1207 SerAlaSerGlyProAlaLeuThrSerValThrProLeu----- 1221
QY 1407 ACATGTGAGTACGCTTCAGCCCGGGGAAAGATGATTTA-----AGTCC 1454
Db 1222 ---AlaProValValProAlaAlaProGlyProProSerLeuGlnProSerGlyAlaSer 1240
QY 1455 CCATTAAGGAAGAGATTGCTCTGACCAAGCGGTGAATGGAGGTTTGGCGAGGACGGCT 1514
Db 1241 ProSerAlaSerAlaLeu-----ThrLeuGlyLeuAlaThrAla 1253
QY 1515 CCAAGATCTGGGTCAATGAGGAGACCAAGCTGGTGTACTTCCAGGACCAAGACAGCC 1574
Db 1254 ProSerLeuSerSerGlnThrProGlyHisProLeuLeuAlaProThrSerSer 1273
QY 1575 -----CCGCTGAGACCAACCTCTACGTGCTGACGTATGAGGCGCGCGCG 1619
Db 1274 HisValProGlyLeuAsnSerThrValAlaProAlaCysSerProValLeuValProAla 1293
QY 1620 AGATGTAACGCTTCACACAGCCCGGCTTCCTCCATAGCTGCTCCATGAGCAGCAACTTCG 1679
Db 1294 SerAlaLeuAlaSerProPherProSerAlaProAsnProAlaPro---AlaGlnAlaSer 1312
QY 1680 ACATGTTGCTGACCACTACAGACAGCTGAGCAGCGCGCTCGCTGACACTTACAAAGC 1739
Db 1313 LeuLeuAlaProAlaSerSerAlaSerGlnAlaLeu-----AlaThrProLeuAlaPro 1330
QY 1740 TGAGCGGCGCCGACGAGCCCTGCAACAGACGCGCGCTTCTGGGCTGACATGATGAG 1799
Db 1331 MetaAlaAlaProGlnThrAlaAlaLeuAlaProSerProAlaProProLeuAlaProLeu 1350
QY 1800 -----AGCAGACAGCTGCCCCCGGATATGATTCCTCCAGATCTTCAT 1847
Db 1351 ProValLeuAlaProSerProGlyAlaAlaProValLeuAlaSerSerGlnThrProVal 1370
QY 1848 -----TCCACAGCGCGCTCGAGTGCCTGCTACGCGATGATTCACA 1889
Db 1371 ProValMetaAlaProSerSerThrProGlyThrSerLeuAlaSerAlaSerProValPro 1390
QY 1890 AGCCCGACGCTTCGACGAGGAAAGAACACCCCA-----CCG 1928
Db 1391 AlaProThrProValLeuAlaProSerSerThrGlnThrMetLeuProAlaProValPro 1410
QY 1929 TCTCTTTGTATGAGAGCCCGCAGGTGAGCTGTGATTAACCTTCCTCAAGGACATCA 1988
Db 1411 SerProLeuProSerProAlaSerThrGlnThrLeuAlaAlaProAlaLeuAlaPro 1430
QY 1989 AGTACTGCGGCTCAACACACTGCTCCCTGGGCTACGCGGTGTTGATTGACGCA 2048
Db 1431 ThrLeuGlyGlySerSer-----ProSerGlnThr-----LeuSerLeuGlyThr 1445
QY 2049 GGGGCTCTGTGACGAGGAGGCTTGCTGGAAGGGGCGCTCAAAAACCAATGGGCCAGG 2108
Db 1446 GlysAsnPro-----GlnGlyProPherProThr----- 1454
QY 2109 TGGAGATGAGAGCAGGTGAGGCGCTGACAGTTGCTGCGAGAGATGAGGCTTCATCG 2168
Db 1455 -----GlnThrLeuSerLeuThrProAlaSerSerLeuValPro 1467
QY 2169 ACCTGAGCCGAGTTGCCATCATGGCT-----GGTCTACGAGGGGCTTCC 2213
Db 1468 ThrProAlaGlnThrLeuSerLeuAlaProGlyProProLeuGlyProThrGlnThrLeu 1487
QY 2214 TCTGCTCATGGGCTAATCCACAAGCCCAAGGTGTTCAGGTGCGCATCGCGGGTGGCC 2273
Db 1488 -----SerLeuAlaProAla-----ProProLeuAlaPro 1497
QY 2274 CGGTACCGCTGTGATGGCTACGACACAGGATACATG----- 2312
Db 1498 AlaSerProValGly---ProAlaProAlaHisThrLeuThrLeuAlaProAlaSerSer 1516
QY 2313 AGCGCTACATGG-----ACGTCCCTGGAACACACAGCAGCGCTATGAGCGGGTT 2363
Db 1517 SerAlaSerLeuLeuAlaProAlaSerValGlnThrLeuThrLeuSerProAlaProVal 1536

```

```

QY 2364 CCG-----TGGCCCTGACGCTGAGAGACTGCCAATGACCCCAACCGCTTGCTATCC 2417
Db 1537 ProThrLeuGlyProAlaAlaAlaGlnThrLeuAlaLeuAlaPro---AlaSerThrGln 1555
QY 2418 TCCAGCGGCTTCCTGAGCAAAACGTGACCTTTTTC----- 2453
Db 1556 SerProAlaSerGlnAlaSerSerLeuValValSerAlaSerGlyAlaAlaProLeuPro 1575
QY 2454 ---ACACAACCTTCCTGCTCTCCCACT-----GATCCGAGCAGGAAACCTTACC 2501
Db 1576 ValThrMetaAlaSer-ArgLeuProValSerLysAspGluProAspThrLeuThrLeuAr 1595
QY 2502 AGCT-----CCAGATCTACCCCAACAGACAGACAGATTCGCTGCGCGAGTGGCG 2555
Db 1595 gSerGlyProProSerProProSerThrAlaThrSerPheGlyProArgProArgAr 1615
QY 2556 A 2556
Db 1615 g 1615

RESULT 15
US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCY
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-09-355-166-1

Alignment Scores:
Pred. No.: 4,15e-11 Length: 657
Score: 249.00 Matches: 152
Percent Similarity: 35.77% Conservative: 112
Best Local Similarity: 20.60% Mismatches: 246
Query Match: 5.19% Indels: 228
DB: Gaps: 34

US-09-976-674-4 (1-2617) x US-09-355-166-1 (1-657)
QY 545 GAATCAAGACCCAGTGTCTGAGGCGC-----CGGATGAGCCCC 583
Db 50 GlnThrLysThrGlnGlySerValProThrPThrHisGlyGluLysArgSerThrAspPro 69
QY 584 AAATCTGCTCCGCGGACCTGCTCTTCTTCCTTCATCAATAC----- 628
Db 70 ArgTrpSerPro---AspGlyArgThrLeuAlaPheIleSerAspArgGluLysPala 88
QY 629 AGCAGCTGTGGGTGGCCACATCGAGACAGGCGAGGAGCGGCGTCACTTGTGCAC 688
Db 89 AlaGlnLeuThrLysIleMetSerThrGlnGlyGluAlaArgLysLeuThrAspIlePro 108
QY 669 CAAGTTATCAATGTCTCTGATGATACCCCAAGTGTGGGGGTGTGCCACTTCCTCATTA 748
Db 109 TyrGlyValSerLysProLeu----- 115
QY 749 CAGGAAGATTGACCGCTTCACTGGTACTGTGTCGCCACACAGCTCCTGGAAGGT 808
Db 116 -----TrpSerPro-----AspGly 120
QY 809 TCAGAGGCGCTTCAGACGCTGCGAATC---CTGTATGAGGAATCGATGATCGAGGTG 865
Db 121 GluSerIleLeuValThrIleSerLeuGlyGluSerIleAspAspArg----- 138

```

```

QY 866 GAGGTCAATTCACGTCCCTCTCTGCGTAGAAGAAAGACGACGCTGATCGGTAC 925
    |||::: |||
Db 139 -----GluysThrGlnGlnAspSerTyrGlu--- 147
QY 926 CCCAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 985
    |||::: |||
Db 148 -----ProValGlnValGlnGlnLeuSerTyrLysArgAsp 159
QY 986 AGCCAGGCGC-----AAGATCGTCGACG 1009
    |||::: |||
Db 160 GYLysGlnLeuThrArgGlnAlaTyrAlaGlnLeuValLeuValSerValLysSerGly 179
QY 1010 CAGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1069
    |||::: |||
Db 180 GluMetLysGlnLeuThrSer-----HisLysAlaAspHis---Gly 192
QY 1070 AGGCGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1129
    |||::: |||
Db 193 AspProAlaPheSerProAspArgLys-----TyrLeuValPhe----- 205
QY 1130 CAGTGGCTCCAGCTCGTCTCTCTCCCGCGCGCTGTTCAATCCGACGACGACGAC 1189
    |||::: |||
Db 206 -----SerAlaAsnLeuThr 210
QY 1190 GACGACGCGGTACGCTCTCTGCGACAGCTGTCCGACGACGACGACGACGACGAC 1249
    |||::: |||
Db 211 GluThrAspAspAlaSer----- 216
QY 1250 GAGGAGGTCAACCAAGCTGGATCATGTTGATGATCTTTATCCCTCCCAATCA 1309
    |||::: |||
Db 217 -----LysProHisAspValTyr----- 222
QY 1310 GAGGAGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1369
    |||::: |||
Db 223 -----IleMetSerLeuGlnSerGlyAspLeuLys----- 232
QY 1370 TTGTACAAATCAACCGCGCTTTTAAATCCAGAGGCTACGATGGAGTACGACGCTTAC 1429
    |||::: |||
Db 233 -----GlnValThrProHisArgGlySerPheGlySerSerSerPheSer 247
QY 1430 CCC-----GGGGAAGATGATTTAATGATCCCATTAAG 1462
    |||::: |||
Db 248 ProAspGlnArgTyrLeuAlaLeuAlaLeuGlnLysGlnLysGlnLys----- 263
QY 1463 GAAGAGATTCGCTGACGACGACGACGACGACGACGACGACGACGACGACGAC 1492
    |||::: |||
Db 264 -----AsnAlaThrLeuSerLysAlaTyrLeuTyrAspIleGlnGlnLysArgLeuThr 281
QY 1493 -----GAGTTTGGCGGAGGAC----- 1510
    |||::: |||
Db 282 CysLeuThrGlnMetLeuAspValHisLeuAlaAspAlaLeuIleGlyAspSerLeuIle 301
QY 1511 ---GGTCCAG-----ATCTGGGTCAATGAGGACGACGACGACGACGACGACGAC 1558
    |||::: |||
Db 302 GlyGlnAlaIleValGlnArgProIleTyr---ThrLysAspSerGlnGlyPheTyrValIle 320
QY 1559 GGCACCAAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1618
    |||::: |||
Db 321 GALTThrAspGlnIle-----SerThrGlnIleTyrTyrIleSerIleGlnGlnLeuVal 338
QY 1619 GAGATGCTACGCTCAACGACGACGACGACGACGACGACGACGACGACGACGAC 1678
    |||::: |||
Db 339 TyrProIleArgLeuGlnLysGln---TyrIleAsnSerPheSerLeuSerProAspGln 357
QY 1679 GACATGTTGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1738
    |||::: |||
Db 358 GlnHisPheIleLeuValSerValThrLysProAspArgProSer-----GluLeuTyrSer 375
QY 1739 CTGAGGCGCGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1798
    |||::: |||
Db 376 Ile-----ProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 389
QY 1799 GAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1858

```

```

Db 390 AspLysPheValArgGlnHisThrIleSerIleProGlnGlnIleGlnTyrAlaThrGlu 409
    |||::: |||
QY 1859 TCGGATGCGCGCGCTCTACGCGATGATCTACAGCCCGACGCGCTTGACGACGACGAC 1918
    |||::: |||
Db 410 AspGlnValMetValAsnGlyTyrPheMetArgProAlaGlnIleGlnGlnGlnThr 429
QY 1919 CACCCGACGCGCTCTTGTATATAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1978
    |||::: |||
Db 430 TyrProIleLeuAsnIleHisGlyGlyProHisMetMetTyrGlnHisThrTyrPhe 449
QY 1979 AAGGACATCAAGTACTTCCGCGCTCAACACACGCGCGCTCTGCGGTACGCGGTGTG 2038
    |||::: |||
Db 450 His-----GluPheGlnValLeuAlaIleValGlnGlyTyrAlaValTyr 464
QY 2039 ATTGACGCGCGCGCTCTCTACGCGCGCGCTTGCGGTCAAGGCGCGCGCGCGCGCGCG 2098
    |||::: |||
Db 465 IleAsnProAlaArgGlnSerHisGlyTyrGlnGlnGlnPheValAsnAlaValArgGlyAsp 484
QY 2099 ATGGCGCGCGGTGAGATCGAGACGACGACGACGACGACGACGACGACGACGACGAC 2158
    |||::: |||
Db 485 TyrGlnGlnLysAspTyrArgAspValMetGlnAlaValAspGlnAlaIleLysArgAsp 504
QY 2159 GCGTTCATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2218
    |||::: |||
Db 505 ProHisIleAspProLysArgLeuGlnValThrGlnGlnGlnGlnGlnGlnGlnGlnGln 524
QY 2219 -----CTCATGGGCGCTATCCACACGACGACGACGACGACGACGACGACGACGAC 2272
    |||::: |||
Db 525 AsnThrIleValGlnGlnThrAsnArg-----PheLysAlaAlaValAlaThrGlnArg 541
QY 2273 CCGGTACGCGCTGATGCGCTAC-----GACACAGGCTAC-----ACTGAG 2314
    |||::: |||
Db 542 SerIleAsnThrIleSerPheHisGlyValSerIleGlyTyrPhePheThrAsp 561
QY 2315 CGCTACATG-----GACGTCCCTGACACACGACGACGACGACGACGACGACGACGAC 2368
    |||::: |||
Db 562 TyrGlnLeuGlnHisAspMetPheGlnAspThrGlnLysLeuThrAspArgSerProLeu 581
QY 2369 -----GCCGTGACGCGGAGGACGACGACGACGACGACGACGACGACGACGACGAC 2422
    |||::: |||
Db 582 LysTyrAlaAlaAsnValGln-----ThrProLeuLeuIleLeuHis 595
QY 2423 GCGTCTCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2482
    |||::: |||
Db 596 GlnGlnArgAspArgCysProIleGlnGlnAlaGlnGlnLeuPheIleAlaLeuLys 615
QY 2483 CGAGCAGGGAACCTTACGACGCTCCAGATCTACCCGACGACGACGACGACGACGACGAC 2536
    |||::: |||
Db 616 LysMetGlnLysGlnThrLysLeuValArgPheProAsnAlaSerHisAsnLeu 633

```

Search completed: December 12, 2002, 11:54:35
 Job time : 63.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 07:18:36 ; Search time 4477 Seconds

(Without alignments)
17011.848 Million cell updates/sec

Title: US-09-976-674-4

Perfect score: 2617

Sequence: 1 caagcttaccatgcccaccca.....tctgagcgcgccgagatccg 2617

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_man:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2592.4	99.1	3716	AX480934	AX480934 Sequence
2	2592.4	99.1	4295	AF452102	AF452102 Homo sapi
3	1854.6	70.9	3243	HSMB05448	AL834376 Homo sapi
4	1477.4	56.5	2261	AK054656	AK054656 Homo sapi
5	1322	50.5	2546	BC000970	BC000970 Homo sapi
6	862.6	33.0	2649	AX354795	AX354795 Sequence
7	862.6	33.0	3106	AX342633	AX342633 Sequence
8	862.6	33.0	3143	AX354793	AX354793 Sequence
9	861	32.9	3127	AF221634	AF221634 Homo sapi
10	651.8	24.9	3130	BC030688	BC030688 Homo sapi
11	573.8	21.9	2510	AX338497	AX338497 Sequence
12	568.2	21.7	2842	AX405770	AX405770 Sequence
13	557.8	21.7	2778	AK000290	AK000290 Homo sapi
14	516.6	19.7	2161	AK027826	AK027826 Homo sapi
15	434.6	16.6	2668	AX405771	AX405771 Sequence
16	371	14.2	382	AX070256	AX070256 Sequence
17	344.2	13.2	1083	AF221637	AF221637 Homo sapi
18	313	12.0	1669	AF221636	AF221636 Homo sapi
19	265	10.1	43501	AC005783	AC005783 Homo sapi
20	212.8	8.1	873	AX083130	AX083130 Sequence
21	212.8	8.1	1197	AF221635	AF221635 Homo sapi
22	192	7.3	22610	AC026385	AC026385 Homo sapi
23	192	7.3	232190	AC074167	AC074167 Mus muscu
24	190	7.3	180238	AC073688	AC073688 Mus muscu
25	183.8	7.0	156324	AC120079	AC120079 Rattus no
26	180	6.9	39514	AC005594	AC005594 Homo sapi
27	161	6.2	12300	AE005887	AE005887 Caulobact
28	151.2	5.8	10866	AE012516	AE012516 Xanthomonas
29	140.4	5.4	2461	DB3263	DB3263 Xanthomonas
30	133.8	5.1	10359	AE012053	AE012053 Xanthomonas
31	126.8	4.8	22030	SC9E12	SC9E12 Streptomy
32	123	4.7	136324	AC120079	AC120079 Rattus no
33	104.8	4.0	10554	AE011686	AE011686 Xanthomon
34	103.4	4.0	59888	AP004212	AP004212 Oryza sat
35	103.4	4.0	218530	AP004168	AP004168 Oryza sat
36	103	3.9	4650	AY122228	AY122228 Drosophila
37	100.8	3.9	20310	AY010120	AY010120 Xanthomon
38	98.2	3.8	12463	AE012059	AE012059 Xanthomon
39	96.4	3.7	10895	AE012522	AE012522 Xanthomon
40	96	3.7	116580	AC112162	AC112162 Mus muscu
41	96	3.7	251899	AC122463	AC122463 Mus muscu
42	90	3.4	147760	AC011846	AC011846 Homo sapi
43	90	3.4	176282	AC068507	AC068507 Homo sapi
44	90	3.4	198295	AC105129	AC105129 Homo sapi
45	88.4	3.4	9984	AE012475	AE012475 Xanthomon

ALIGNMENTS

RESULT 1

AX480934

LOCUS

DEFINITION

AX480934 Sequence 30 from Patent WO0246383.

ACCESSION

AX480934

VERSION

AX480934.1 GI:22217573

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1

Yue,H., Azimzal,Y., Kallick,D.A., Baughn,M.R., Griffin,J.A., Swarnakar,A., Lal,P.G., Walla,N.K., Hafalla,A.J., Gandhi,A.R., Au-Young,J., Elliott,V.S., Ramkumar,J., Thangavelu,K., Lu,Y.,

Pred. No. is the number of results predicted by chance to have a

Warren, B.A., Lu, D.A., Lee, E.A., Tribouley, C.M., Arvizu, C.,
Deleane, A.M., Yao, M.G., Khan, F.A. and Sanjanwalla, M.M.
Protein modification and maintenance molecules
Patent: WO 0246383-A 30.13-JUN-2002;
Incyte Genomics, Inc. (US)

FEATURES
source
1. 3716
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 7972712CB1"

BASE COUNT 768 a 1174 c 1075 g 699 t
ORIGIN

Query Match 99.1%; Score 2592.4; DB 6; Length 3716;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2593; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 ATGGCCACACCGGGAGCCCAACGCGCGAGCGAGCGAGCGCGCGACAGATGACCG 70
DB 330 ATGGCCACACCGGGAGCCCAACGCGCGAGCGAGCGCGCGCGACAGATGACCG 389
QY 71 GCGCGCGCTTCAGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 130
DB 390 GCGCGCGCTTCAGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 449
QY 131 AGCGCAGACTCGGCGCTCATTTGTCAACAGCGCGCGCGCGCGAGCTTCAGTTGTG 190
DB 450 AGCGCAGACTCGGCGCTCATTTGTCAACAGCGCGCGCGCGCGAGCTTCAGTTGTG 509
QY 191 AAGACGAGTGTGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 250
DB 510 AAGACGAGTGTGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 569
QY 251 AGCGGAGAGTCCCTCTACTCTGAGATGCCAAGAGGCGGGAAGAGGCTGTG 310
DB 570 AGCGGAGAGTCCCTCTACTCTGAGATGCCAAGAGGCGGGAAGAGGCTGTG 629
QY 311 CTGCTCTGCTGAGAGAGAGTGTGATCATTTCCAGGCGCGCGCGCGCGCGCGCGCG 370
DB 630 CTGCTCTGCTGAGAGAGAGTGTGATCATTTCCAGGCGCGCGCGCGCGCGCGCGCG 689
QY 371 TACTCTGAGAGAGAGAGTGTGAGGAGCGGAACGCGCTGGGGTCTTCCGATACG 430
DB 690 TACTCTGAGAGAGAGAGTGTGAGGAGCGGAACGCGCTGGGGTCTTCCGATACG 749
QY 431 TCTAGACTTCCACAGCGAGTGTGCTCTCTCCAGGCGCGCGCGCGCGCGCGCGCG 490
DB 750 TCTAGACTTCCACAGCGAGTGTGCTCTCTCCAGGCGCGCGCGCGCGCGCGCGCG 809
QY 491 CACTGTGCGAGCGCGCGCAAGAGCGCTTCATGCTGCCCTATGAACGCGTGAATC 550
DB 810 CACTGTGCGAGCGCGCGCAAGAGCGCTTCATGCTGCCCTATGAACGCGTGAATC 869
QY 551 AAGACCGAGTGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 610
DB 870 AAGACCGAGTGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 929
QY 611 TTCTCTTCATCATACAGCGAGCTGTGGTGGCAACATCGAGAGCGAGCGAGCGAG 670
DB 930 TTCTCTTCATCATACAGCGAGCTGTGGTGGCAACATCGAGAGCGAGCGAGCGAG 989
QY 671 CGGCTGACCTTCGCGCAAGGTTTATCCATGTCTGTGATGACCCCAAGTGTGCGGG 730
DB 990 CGGCTGACCTTCGCGCAAGGTTTATCCATGTCTGTGATGACCCCAAGTGTGCGGG 1049
QY 731 GTGGCCACCTTCTGATACAGAGAGTGTGACCGCTTCAGTGGGTACTGGTGGCCCC 790
DB 1050 GTGGCCACCTTCTGATACAGAGAGTGTGACCGCTTCAGTGGGTACTGGTGGCCCC 1109
QY 791 ACAGCGCTTCGGGAAGGTTGAGAGGCGCTCAAGAGCGCTGGAATCTGTATGAGGAATC 850
DB 1110 ACAGCGCTTCGGGAAGGTTGAGAGGCGCTCAAGAGCGCTGGAATCTGTATGAGGAATC 1169

QY 851 GATGAGTCCGAGGTGAGGTGATTCACGTCCTCTCTGCTGAGAGAGAGAGACG 910
DB 1170 GATGAGTCCGAGGTGAGGTGATTCACGTCCTCTCTGCTGAGAGAGAGAGACG 1229
QY 911 GATCTGATCGGTACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 970
DB 1230 GATCTGATCGGTACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1289
QY 971 GAGTTCAGACTGACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1030
DB 1290 GAGTTCAGACTGACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1349
QY 1031 CCTTCAGCTCGCTGTTCGCCAGAGGTGAGTACATCCGAGAGAGAGAGAGAGAGAG 1090
DB 1350 CCTTCAGCTCGCTGTTCGCCAGAGGTGAGTACATCCGAGAGAGAGAGAGAGAGAG 1409
QY 1091 GGCAGATACGCTGGGCGAGTTCCTGAGCCGCGCGCGCGCGCGCGCGCGCGCGCG 1150
DB 1410 GGCAGATACGCTGGGCGAGTTCCTGAGCCGCGCGCGCGCGCGCGCGCGCGCGCG 1469
QY 1151 CTGCGCGCGCGCGCTTCATCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1210
DB 1470 CTGCGCGCGCGCGCTTCATCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1529
QY 1211 AGAGCTGTCCAGAGATGTCCAGCCGTATGTGTGACAGAGAGAGTCCACCAAGTGTG 1270
DB 1530 AGAGCTGTCCAGAGATGTCCAGCCGTATGTGTGACAGAGAGAGAGTCCACCAAGTGTG 1589
QY 1271 ATCAATTTATGATCTCTCTATCCCTTCCCAATCAGAGAGAGAGAGAGAGAGTGTG 1330
DB 1590 ATCAATTTATGATCTCTCTATCCCTTCCCAATCAGAGAGAGAGAGAGAGTGTG 1649
QY 1331 TTTCTCGCGCGCAATGATGATCAAGACGCGCTTGCCATTGTGACAAAGTCACCGCGTT 1390
DB 1650 TTTCTCGCGCGCAATGATGATCAAGACGCGCTTGCCATTGTGACAAAGTCACCGCGTT 1709
QY 1391 TTAAGATCCAGAGGCTAGATTTGAGTGAAGCCCTTACGCGCGCGCGAGAGATGAATTTAAG 1450
DB 1710 TTAAGATCCAGAGGCTAGATTTGAGTGAAGCCCTTACGCGCGCGCGAGAGATGAATTTAAG 1769
QY 1451 TGGCCCTTAAAGAGAGATGCTGTGACAGCGGTAAAGAGGTTTGGAGAGCGAC 1510
DB 1770 TGGCCCTTAAAGAGAGATGCTGTGACAGCGGTAAAGAGGTTTGGAGAGCGAC 1829
QY 1511 GGCCTCAAGATGTGGGTCAATGAGAGAGACCAAGTGTGTACTTCCAGGCGCACCAAGAGAC 1570
DB 1830 GGCCTCAAGATGTGGGTCAATGAGAGAGACCAAGTGTGTACTTCCAGGCGCACCAAGAGAC 1889
QY 1571 ACGCGCTGAGACACACCTCTACGTGTGATGAGGCGCGCGCGCGAGATGTGACG 1630
DB 1890 ACGCGCTGAGACACACCTCTACGTGTGATGAGGCGCGCGCGCGAGATGTGACG 1949
QY 1631 CTCACACGCGCGCGCTTCTCCATAGCTGTGCTCAATGAGCGCAAGACTTCGATGTGTC 1690
DB 1950 CTCACACGCGCGCGCTTCTCCATAGCTGTGCTCAATGAGCGCAAGACTTCGATGTGTC 2009
QY 1691 AGCCACTACAGACCGCTGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1750
DB 2010 AGCCACTACAGACCGCTGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2069
QY 1751 GAGCAGACCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1810
DB 2070 GAGCAGACCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2129
QY 1811 TGGCGCGCGGATTAATGTCTTCCAGAGATCTTCATTTCCAGAGCGCGCTGGATGTGCGG 1870
DB 2130 TGGCGCGCGGATTAATGTCTTCCAGAGATCTTCATTTCCAGAGCGCGCTGGATGTGCGG 2189
QY 1871 CTCTACGAGATGATCTCAAG 1930
DB 2190 CTCTACGAGATGATCTCAAG 2249
QY 1931 CTCTTTGTATATGAGAGCGCGCGAGGTGACGTGTGTAATCTTCAAGAGCATCAAG 1990


```

|||||
Db 2250 CTCTTTATATGAGGAGCCGCCAGGTGCTGTAATACTCTCAAGCATCAG 2309
Qy 1991 TACTTGGGGTCAACACACTGGCTCCCTGGGCTACGCCCGGTGGATTGAGGCAAG 2050
Db 2310 TACTTGGGGTCAACACACTGGCTCCCTGGGCTACGCCCGGTGGATTGAGGCAAG 2369
Qy 2051 GGCTCTGTGACGAGGAGGCTGGTGAAGGGGCGCTGAAGAAACCAATGGCCAGGTG 2110
Db 2370 GGCTCTGTGACGAGGAGGCTGGTGAAGGGGCGCTGAAGAAACCAATGGCCAGGTG 2429
Qy 2111 GAGATGAGAGACACAGGTGAGAGGCGCTGAGTTCGTGGCCGGAATATGAGCTTCATCGAC 2170
Db 2430 GAGATGAGAGACACAGGTGAGAGGCGCTGAGTTCGTGGCCGGAATATGAGCTTCATCGAC 2489
Qy 2171 CTGAGCGGAGTGGCTCATCATGCTGCTCTACGAGGGGCTTCTCTGCTATGAGGCTA 2230
Db 2490 CTGAGCGGAGTGGCTCATCATGCTGCTCTACGAGGGGCTTCTCTGCTATGAGGCTA 2549
Qy 2231 ATCCACAAGCCCCAGTGTTCAGTGGCCATCGCGGGGTGCCCGGTCACCGTGGATG 2290
Db 2550 ATCCACAAGCCCCAGTGTTCAGTGGCCATCGCGGGGTGCCCGGTCACCGTGGATG 2609
Qy 2291 GCCTACGACACAGGGTACACTGAGCGCTACATGAGACGTCCCTGAGAACACACAGCAGCGC 2350
Db 2610 GCCTACGACACAGGGTACACTGAGCGCTACATGAGACGTCCCTGAGAACACACAGCAGCGC 2669
Qy 2351 TATGAGCGGGGTTCCGTGGCGCTGACGTGGAGAAAGTCCCAATGAGCCCAACCGCTTG 2410
Db 2670 TATGAGCGGGGTTCCGTGGCGCTGACGTGGAGAAAGTCCCAATGAGCCCAACCGCTTG 2729
Qy 2411 CTATTCCTCCACGGCTTCTCTGAGAAAGCTGACATTTTCCACACAACCTCTCTGTC 2470
Db 2730 CTATTCCTCCACGGCTTCTCTGAGAAAGCTGACATTTTCCACACAACCTCTCTGTC 2789
Qy 2471 TCCCACTGATCCGAGGAGGAAACCTTACAGCTCCAGATCTACCCCAACAGAGACAC 2530
Db 2790 TCCCACTGATCCGAGGAGGAAACCTTACAGCTCCAGATCTACCCCAACAGAGACAC 2849
Qy 2531 AGTATTTGGCTGCCCCGAGTGGGGGAGGACATATGAAGTACAGCTTCTCTACAG 2590
Db 2850 AGTATTTGGCTGCCCCGAGTGGGGGAGGACATATGAAGTACAGCTTCTCTACAG 2909
Qy 2591 GAATACCTCTGAGC 2604
Db 2910 GAATACCTCTGAGC 2923

RESULT 2
AF452102 4295 bp mRNA linear PRI 08-JUL-2002
LOCUS Homo sapiens dipeptidyl peptidase-like protein 9 mRNA, complete
DEFINITION cds.
ACCESSION AF452102
VERSION AF452102.2 GI:21704252
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4295)
AUTHORS Olsen, C. and Magtman, N.
TITLE Identification and characterization of a novel member of the
dipeptidyl peptidase IV-related family
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4295)
AUTHORS Olsen, C. and Magtman, N.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-2001) Biotechnology, Novo Nordisk A/S, Novo Alle,
Bagsvaerd DK-2880, Denmark
REFERENCE 3 (bases 1 to 4295)
AUTHORS Olsen, C. and Magtman, N.
TITLE Direct Submission

```

```

JOURNAL Submitted (08-JUL-2002) Biotechnology, Novo Nordisk A/S, Novo Alle,
REMARK Bagsvaerd DK-2880, Denmark
COMMENT Sequence update by submitter
FEATURES On Jul 8, 2002 this sequence version replaced gi:17865310.
SOURCE Location/Qualifiers
1..4295
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
366..4295
/note="DPP9; peptidase family 9B; LOC91039; HS.237617"
/codon_start=1
/product="dipeptidyl peptidase-like protein 9"
/protein_id="AA147179.1"
/db_xref="GI:17865311"
/translation="MATTPPTDRGDAAATDDPARFOYOKHSMGLRSIIHGRKY
SGILVNKAPHDFOFVOKTDEGSHSLHYLYGMFGSRENSLYSEIPIKVRKALIL
LSMKOMLDFHOAPFHGVSREELERRLGFGVITSDYDFHESGLFLFOANSIF
HCRDGGKNGFMVSPMKLEIKTOSGSRMDPKICPADAPAFSPFTNNSDLVANTGE
ERRITPCGSLSNVLDPKSAGVATPIYIOEBDFRTYMMCTPASMGSGLKTRIL
YEYDESEVEYIHYPSALBERKTDSTRIYRTGSKNPKIALKAEFOYDSOGKIVSTO
EKELVQPFSSLPKVEYIARAGWTRDKRYAMAMFLDRQOMLQVLPLPALTPTSTEN
EEOFLASARAVPRNVQPYVYEEVTNWIMVHDIFFPPQSEGEDELCTFRANCKTG
FCHLYKTVAILKSGYDMSEPFSGDEDFCPKIEELIATSGEVEYLARHSGKIWNE
ETKLYVRQGRKIDPLRHLHYVEYEAAGELVRLTTPERSHSCSMOYDFMFSYSYV
STPVCVHYVRLSDPDDPLHKOPRFNASHMEBAASCPRDYIPPELTHRTSDVRLIY
IYRPHALQPKRHEPTLVLYVGPOVOLVNNSPFGIKRLYRLNTLASLYAAVVIDGRS
CORLRFEGALIKQMGVEIEDOVEGLQFAEKYGFIDLSRYALHSGSYGFLSLMGL
IHKPYEKFVALIAGAPYVMAAYDTGERMYADPENNOHSEYAGSVLAHYEKLPEPN
RLILHGFLEDNVHFHTNFLVSQLINAGRPYQIYIPNERHRSIRCESEBHEVTLL
HFIQEVY"
BASE COUNT 922 a 1346 c 1220 g 807 t
ORIGIN
Query Match 99.1%; Score 2592.4; DB 9; Length 4295;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2593; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 11 ATGGCCACCACCGGAGCCCAACGCGGACGAGCGACGCCACAGATGACCG 70
Db 366 ATGGCCACCACCGGAGCCCAACGCGGACGAGCGAGCGCCACAGATGACCG 425
Qy 71 GCCGCGCGCTTCAGGTGACAGACACTGTTGGGAGCGGCTCCGAGATCATCCAGGC 130
Db 426 GCCGCGCGCTTCAGGTGACAGACACTGTTGGGAGCGGCTCCGAGATCATCCAGGC 485
Qy 131 AGCCGCAAGTACTGGGCTCATTTGTCAACAAGGCGCCCAAGACTTCAGATTGTGCGAG 190
Db 486 AGCCGCAAGTACTGGGCTCATTTGTCAACAAGGCGCCCAAGACTTCAGATTGTGCGAG 545
Qy 426 GCCGCGCGCTTCAGGTGACAGACACTGTTGGGAGCGGCTCCGAGATCATCCAGGC 485
Db 426 GCCGCGCGCTTCAGGTGACAGACACTGTTGGGAGCGGCTCCGAGATCATCCAGGC 485
Qy 191 AAGAGGATGAGTGTGGGCCCCACTCCACCGCTCTACTACCTGGAATGCCATATGCG 250
Db 546 AAGAGGATGAGTGTGGGCCCCACTCCACCGCTCTACTACCTGGAATGCCATATGCG 605
Qy 251 AGCGGAGAACTCCCTCTACTCTGAGANTTCCCAAGAGTCCGGAAGAGGCTCTG 310
Db 606 AGCGGAGAACTCCCTCTACTCTGAGANTTCCCAAGAGTCCGGAAGAGGCTCTG 665
Qy 311 CTGCTCTCTCTGGAACAGATGCTGATATTTCCAGGCGACGCCACCATGGGGTGC 370
Db 666 CTGCTCTCTCTGGAACAGATGCTGATATTTCCAGGCGACGCCACCATGGGGTGC 725
Qy 371 TACTCTGGGAGGAGAGCTGTGAGGAGCGGAAAGCTGTGGGGTCTTGGCATCAGC 430
Db 726 TACTCTGGGAGGAGAGCTGTGAGGAGCGGAAAGCTGTGGGGTCTTGGCATCAGC 785
Qy 431 TCTACGACTTCCACAGGAGAGTGGCTCTTCTCTTCCAGGCCACAGAACGCTCTTC 490
Db 786 TCTACGACTTCCACAGGAGAGTGGCTCTTCTCTTCCAGGCCACAGAACGCTCTTC 845
Qy 491 CACTGTGCGAGCGGCGGAAGAGCGCTTCAATGATGCTCCCTATGAACCGCTGGAATTC 550

```

Db 846 CACTGCCGCGAGCGCGGCAAGAACGGCTTCAATGCTGTCCTCATATGAACCGCTGGAATC 905
QY 551 AAGACCCAGTGTACAGGGCCCGGATGGACCCCAAAATGTGCCCTGGCGACCGCTTGC 610
Db 906 AAGACCCAGTGTACAGGGCCCGGATGGACCCCAAAATGTGCCCTGGCGACCGCTTGC 965
QY 611 TTCTCTTCATCATATAACAGCGACCTGTGGGTGGCCAAATCGAGACAGGCGAGGACGG 670
Db 966 TTCTCTTCATCATATAACAGCGACCTGTGGGTGGCCAAATCGAGACAGGCGAGGACGG 1025
QY 671 CGGCTGACCTTCTGCCCAAGGTTTATCCAAATGCTGGATGACCCCAAGTGTGGGGT 730
Db 1026 CGGCTGACCTTCTGCCCAAGGTTTATCCAAATGCTGGATGACCCCAAGTGTGGGGT 1085
QY 731 GTGGCCACCTTCTGTCATACAGGAAGATTGACCGGCTTCACTGGGATCTGTGGTGGCCC 790
Db 1086 GTGGCCACCTTCTGTCATACAGGAAGATTGACCGGCTTCACTGGGATCTGTGGTGGCCC 1145
QY 791 ACAGCCTCTGGGAAGGTTGAGAGGGCCTGAAGACGCTGCAATCTGTATGAGAAATC 850
Db 1146 ACAGCCTCTGGGAAGGTTGAGAGGGCCTGAAGACGCTGCAATCTGTATGAGAAATC 1205
QY 851 GATGAGTCGAGGTGGAGGTATTCACGTCCCTCTGCTGCGTGAAGAAAGAAAGACG 910
Db 1206 GATGAGTCGAGGTGGAGGTATTCACGTCCCTCTGCTGCGTGAAGAAAGAAAGACG 1265
QY 911 GACTGTATCGGTACCCACAGACAGCGACAGAAATCCCAAGATTGCTTGAACCTGGCT 970
Db 1266 GACTGTATCGGTATCCACAGACAGCGACAGAAATCCCAAGATTGCTTGAACCTGGCT 1325
QY 971 GACTTCACACTGACAGCCAGGCGAGAGATGCTCTGACCCAGGAGAAAGAGCTGTGACG 1030
Db 1326 GACTTCACACTGACAGCCAGGCGAGAGATGCTCTGACCCAGGAGAAAGAGCTGTGACG 1385
QY 1031 CCGTTCAGCTCGGCTTCCGGAAGGTGAGTATACATGCGCAAGGGCGGGTGGACCCGGGAT 1090
Db 1386 CCGTTCAGCTCGGCTTCCGGAAGGTGAGTATACATGCGCAAGGGCGGGTGGACCCGGGAT 1445
QY 1091 GGCAGATACGCTGTGGGCGCATGTTCTGTGACCGGCGGACAGTGGCTCCAGCTGTCTC 1150
Db 1446 GGCAGATACGCTGTGGGCGCATGTTCTGTGACCGGCGGACAGTGGCTCCAGCTGTCTC 1505
QY 1151 CTGCCCCCGGCGCTTTCATCCGAGACAGAGATGAGAGACAGCGGCTTACGCTTGGC 1210
Db 1506 CTGCCCCCGGCGCTTTCATCCGAGACAGAGATGAGAGACAGCGGCTTACGCTTGGC 1565
QY 1211 AGAGCTGTCCCGAGATGTCCAGCGCTATGTGTGTGACGAGGAGGTCAACAAGTGTGG 1270
Db 1566 AGAGCTGTCCCGAGATGTCCAGCGCTATGTGTGTGACGAGGAGGTCAACAAGTGTGG 1625
QY 1271 ATCAATGTTCATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGACGCTTGC 1330
Db 1626 ATCAATGTTCATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGACGCTTGC 1385
QY 1331 TTCTCTCGGCGCATGATGACAGACCGGCTTCTGCCATTGTGACAAAGTCAACCGCGTT 1390
Db 1686 TTCTCTCGGCGCATGATGACAGACCGGCTTCTGCCATTGTGACAAAGTCAACCGCGTT 1745
QY 1391 TTAAATATCCAGGGCTACGATTTGAGTGGAGCGCTTCAAGCCCGGGGAAGATGATTTAG 1450
Db 1746 TTAAATATCCAGGGCTACGATTTGAGTGGAGCGCTTCAAGCCCGGGGAAGATGATTTAG 1805
QY 1451 TGCCCATTTAAGGAAGAGATTTGCTGTGACAGCGGTGAATGGAGGTTTGGGAGGAC 1510
Db 1806 TGCCCATTTAAGGAAGAGATTTGCTGTGACAGCGGTGAATGGAGGTTTGGGAGGAC 1865
QY 1511 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGACTTCCAGGGCACCAGAGAC 1570
Db 1866 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGACTTCCAGGGCACCAGAGAC 1925
QY 1571 ACGCCCTGGAGACCAACCTCTACGTGTGAGTATGAGGCGGCGGAGATCGTACGC 1630
Db 1926 ACGCCCTGGAGACCAACCTCTCTACGTGTGAGTATGAGGCGGCGGAGATCGTACGC 1985

QY 1631 CTCACACGCCCGGCTTCTCCCATAGCTGTGCTCATGAGCCAGAATCTGACATTTGCTC 1690
Db 1986 CTCACACGCCCGGCTTCTCCCATAGCTGTGCTCATGAGCCAGAATCTGACATTTGCTC 2045
QY 1691 AGCCACTACAGACAGCTGTGACAGCGCGCTTGGCTGACGTTTCAAGAGTGTAGCGGCC 1750
Db 2046 AGCCACTACAGACAGCTGTGACAGCGCGCTTGGCTGACGTTTCAAGAGTGTAGCGGCC 2105
QY 1751 GAGCAGACCCCTGTCACAGAGCCCGGCTTGGGCTGAGCATGATGAGGAGCGCCAGC 1810
Db 2106 GAGCAGACCCCTGTCACAGAGCCCGGCTTGGGCTGAGCATGATGAGGAGCGCCAGC 2165
QY 1811 TGCCCCCGGATTTATGTTCTCCAGAGATCTTCATTTCCACAGCGGCTGAGTGTGCG 1870
Db 2166 TGCCCCCGGATTTATGTTCTCCAGAGATCTTCATTTCCACAGCGGCTGAGTGTGCG 2225
QY 1871 CTCTACGGCATGATCTAAGACCCACGCTTGCAGCGCAGGGAAGAACACCCACGTC 1930
Db 2226 CTCTACGGCATGATCTAAGACCCACGCTTGCAGCGCAGGGAAGAACACCCACGTC 2285
QY 1931 CTCTTGTATATGAGGCGCCCGAGGTGACAGTGTGTAATACCTTCAAGGATCAAG 1990
Db 2286 CTCTTGTATATGAGGCGCCCGAGGTGACAGTGTGTAATACCTTCAAGGATCAAG 2345
QY 1991 TACTTGGGCTCAACACACTGTGCTCCCTGCGCTACGCGCTGTGTGTGATGACGCGAG 2050
Db 2346 TACTTGGGCTCAACACACTGTGCTCCCTGCGGCTACGCGCTGTGTGTGATGACGCGAG 2405
QY 2051 GGCCTCTGTACAGGAGGCTTCCGTTGAAAGGGCCCTGAAAAACCAATGCGCAGGTG 2110
Db 2406 GGCCTCTGTACAGGAGGCTTCCGTTGAAAGGGCCCTGAAAAACCAATGCGCAGGTG 2465
QY 2111 GAGATGAGGACCAAGGTGGAGGGGCTGACAGTGTGCTGCGAGAAATGTGGCTTCATGAC 2170
Db 2466 GAGATGAGGACCAAGGTGGAGGGGCTGACAGTGTGCTGCGAGAAATGTGGCTTCATGAC 2525
QY 2171 CTGAGCGAGTGTGCATCATGTGCTGTCTACGGGGCTTCTCTGCTCATGAGGGCTA 2230
Db 2526 CTGAGCGAGTGTGCATCATGTGCTGTCTACGGGGCTTCTCTGCTCATGAGGGCTA 2585
QY 2231 ATCCACAGCCCCAGGTGTTCAGAGTGGCCATGCGGGGTCCCGGCTACCGTCTGATG 2290
Db 2586 ATCCACAGCCCCAGGTGTTCAGAGTGGCCATGCGGGGTCCCGGCTACCGTCTGATG 2645
QY 2291 GCGTACGACAGGGGTACACTGTGACGCTGACATGAGACGCTCCCTGAGAACACGACGGC 2350
Db 2646 GCGTACGACAGGGGTACACTGTGACGCTGACATGAGACGCTCCCTGAGAACACGACGGC 2705
QY 2351 TATGAGGCGGGTTCGTTGCGCTGTGACAGTGAAGCTGCCCAATGAGCCCAACCGTTG 2410
Db 2706 TATGAGGCGGGTTCGTTGCGGCTGTGACAGTGAAGCTGCCCAATGAGCCCAACCGTTG 2765
QY 2411 CTATATCTCCAGGGCTTCTGAGAGGAAAGCTGACACTTTTCCACACAACTTCTGTC 2470
Db 2766 CTATATCTCCAGGGCTTCTGAGAGGAAAGCTGACACTTTTCCACACAACTTCTGTC 2825
QY 2471 TCCCAACTGATCGAGAGGAAACCTTACAGCTCCAGATCTAACCCACGAGAGACAC 2530
Db 2826 TCCCAACTGATCGAGAGGAAACCTTACAGCTCCAGATCTAACCCACGAGAGACAC 2885
QY 2531 AGTATTCGTGCCCCGAGTGGGCGAGCACTATGAAGTCAAGTGTGCTGACTTTTACAG 2590
Db 2886 AGTATTCGTGCCCCGAGTGGGCGAGCACTATGAAGTCAAGTGTGCTGACTTTTACAG 2945
QY 2591 GAATACCTGTGAGC 2604
Db 2946 GAATACCTGTGAGC 2959

RESULT 3
HSM805448
LOCUS HSM805448 3243 bp mRNA linear PRI 12-JUL-2002

```

DEFINITION Homo sapiens mRNA; cDNA DKFZp762M2413 (from clone DKFZp762M2413).
ACCESSION AL834376
VERSION AL834376.1 GI:21740044
KEYWORDS
SOURCE
ORGANISM human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 3243)
REFERENCE Ansoorge, W., Winkner, U., Mewes, H.W., Well, B. and Wiemann, S.
AUTHORS Direct Submission
TITLE Submitted (09-JUL-2002) 1, D-85764 Neuberberg, GERMANY
JOURNAL Clome from S. Wiemann, Molecular Genome Analysis, German Cancer
COMMENT Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp762M2413) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES
    source
        1..3243
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /map="19p13.3"
        /clone="DKFZp762M2413"
        /tissue-type="melanoma (Memo cell line)"
        /clone_lib="762 (synonym: hmel2). Vector pSport1; host
        DH10B; sites NotI + SalI"
        /dev_stage="adult"
        1..3243
        /gene="DKFZp762M2413"
        <1..1887
        /gene="DKFZp762M2413"
        /note="unknown protein"
        /codon_start=1
        /product="hypothetical protein"
        /protein_id="CAD39039.1"
        /db_xref="GI:21740045"
        /translation="1MWANIETGEERRLTFCHGLSNVLDDPKSAGVAFVIOEEDR
        FTGYWPCPTAEMSGSEGLKTLRILEYEVDESEVEVHPSPALEEKRTDYSRPRGS
        KNKRIALKEFOTDSOGKIVSTOEKEIVQESLPKVEYIRAKRTGRKXAMAF
        LRPQOMLOLVLPALFLPTSPNEBORLASAAVARNQPVYVEVYVWVNIYNDI
        FYFPNORERTSASFAPMARPAALICTKSPFLKSQGDMEPSYEDGDEKCPK
        EETALTSGEWEVLARHGSKIWNEETKLYFQGTQDTPLEHLLVVSYEAAEIVRLT
        TPQFSHSCMSQNFDMKSVSHYSSTPCVHVYKLSGPDDELPHKQREVASMMKESAS
        CPDYPPEIFHFHTSDVLYGMITYKPHALOPGRKHPYLVFYGGPOVOLVNSFEG
        IKYLRNLTASIGYAVVVIDGSGCORGRFEGALKNOMQVEIEOVDELOFVAEKY
        GFTDSLRAVIGHGYSYGFSLIMGLIKHPOVEKYAICAPVTVMAAYDCTERYMNP
        ENNOHCYBAGSVALHVEKLPNEPNRLILHGFIDENVHFFHTNLFVLSOLIRAKPYOL
        QEYL"
        3199..3204
        /gene="DKFZp762M2413"
        3226
        /gene="DKFZp762M2413"
    polyA_signal
        3199..3204
    polyA_site
        /gene="DKFZp762M2413"
    BASE COUNT 731 a 977 c 911 g 624 t
    ORIGIN
        Query Match 70.9%; Score 1854.6; DB 9; Length 3243;
        Best Local Similarity 99.7%; Pred. No. 4.6e-306;
        Matches 1879; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
    QY 635 CTGTGGTGGCCAAATCGAGACAGGCGGAGAGCGGCGCTGACCTTCTGCCCAAGGT 694
        |||||||
        Db 1 CTGTGGTGGCCAAATCGAGACAGGCGGAGAGCGGCGCTGACCTTCTGCCCAAGGT 60
    QY 695 TTATCCAAATGCTGATGACCCCAAGTCTGCGGGTGGCGACACTTCGTCATACAGAA 754
        |||||||
        Db 61 TTATCCAAATGCTGATGACCCCAAGTCTGCGGGTGGCGACACTTCGTCATACAGAA 120
    QY 755 GAGTTGACCGCTTCACTGGGTACTGGTGTGCCCCACACAGCTCTGGGAAGTTCCAGAG 814

```

```

|||
Db 121 GAGTTGACCGCTTCACTGGGTACTGGTGTGCCCCACACAGCTCTGGGAAGTTCCAGAG 180
QY 815 GGCCTCAAGACGCTGCGAATCTCTATGAGAACTCGATGATGCCAGAGTGGAGTCATT 874
|||
Db 181 GGCCTCAAGACGCTGCGAATCTCTATGAGAACTCGATGATGCCAGAGTGGAGTCATT 240
QY 875 CAGTCCCTCTCTGGGCTAGGAAGAAAGAGAGGAGCTGTGCTGATGCCAGAGCA 934
|||
Db 241 CAGTCCCTCTCTGGGCTAGGAAGAAAGAGAGGAGCTGTGCTGATGCCAGAGCA 300
QY 935 GGCAGCAAGATCCCAAGATTGCTTGAATGCTGCTGATGCCAGAGTGGAGTGGAGTGG 994
|||
Db 301 GGCAGCAAGATCCCAAGATTGCTTGAATGCTGCTGATGCCAGAGTGGAGTGGAGTGG 360
QY 995 AAGATGCTCTGACCCAGAGAGAGAGAGAGTGTGTCAGCCCTTCAGCTGCTGCTCCG 1054
|||
Db 361 AAGATGCTCTGACCCAGAGAGAGAGAGAGTGTGTCAGCCCTTCAGCTGCTGCTCCG 420
QY 1055 GTGAGTACATCGCCAGAGGCGGGTGGACCGGGATGGCAATAGCGCTGGGCGCATGTC 1114
|||
Db 421 GTGAGTACATCGCCAGAGGCGGGTGGACCGGGATGGCAATAGCGCTGGGCGCATGTC 480
QY 1115 CTGAGACCGGCCCCAGACAGTGGCTCCAGCTGCTCCCTCCCGCGGCGCTTTCATCCG 1174
|||
Db 481 CTGAGACCGGCCCCAGACAGTGGCTCCAGCTGCTCCCTCCCGCGGCGCTTTCATCCG 540
QY 1175 AGCAGAGAAATGAGAGAGACAGAGAGTGCCTCTGCTCCAGAGTGTGCCAGAAATGTC 1234
|||
Db 541 AGCAGAGAAATGAGAGAGAGAGAGTGCCTCTGCTCCAGAGTGTGCCAGAAATGTC 600
QY 1235 CCGTATGCTGCTACGAGAGAGTGCACCAAGCTCTGATGATGATGATGATGATGATGAT 1294
|||
Db 601 CCGTATGCTGCTACGAGAGAGTGCACCAAGCTCTGATGATGATGATGATGATGATGAT 660
QY 1295 CCGTATGCTGCTACGAGAGAGTGCACCAAGCTCTGATGATGATGATGATGATGATGAT 1354
|||
Db 661 CCGTATGCTGCTACGAGAGAGTGCACCAAGCTCTGATGATGATGATGATGATGATGAT 719
QY 1355 ACCGGCTTCTGCTACGAGAGAGTGCACCAAGCTCTGATGATGATGATGATGATGATGAT 1413
|||
Db 720 ACCGGCTTCTGCTACGAGAGAGTGCACCAAGCTCTGATGATGATGATGATGATGATGAT 779
QY 1414 GAGTACCGCTTCTGCTACGAGAGAGTGCACCAAGCTCTGATGATGATGATGATGATGATGAT 1473
|||
Db 780 GAGTACCGCTTCTGCTACGAGAGAGTGCACCAAGCTCTGATGATGATGATGATGATGATGAT 839
QY 1474 TCTGACGAGCGGTGAATGAGAGTGTGTTGGCGAGGACAGGCTCCAAAGATCTGGTCAATGA 1533
|||
Db 840 TCTGACGAGCGGTGAATGAGAGTGTGTTGGCGAGGACAGGCTCCAAAGATCTGGTCAATGA 899
QY 1534 GGAGACCAAGCTGTGTACTTCCAGAGGACCAAGAGACAGCGCGTGGAGACCAACTCTA 1593
|||
Db 900 GGAGACCAAGCTGTGTACTTCCAGAGGACCAAGAGACAGCGCGTGGAGACCAACTCTA 959
QY 1594 CCGTGTGAGCTATGAGAGCGGCGGCGGAGATGATGATGATGATGATGATGATGATGATGAT 1653
|||
Db 960 CCGTGTGAGCTATGAGAGCGGCGGCGGAGATGATGATGATGATGATGATGATGATGATGAT 1019
QY 1654 TAGTGTCTCATGACGACCAAGCTTGCATGATGATGATGATGATGATGATGATGATGATGAT 1713
|||
Db 1020 TAGTGTCTCATGACGACCAAGCTTGCATGATGATGATGATGATGATGATGATGATGATGAT 1079
QY 1714 GCGGCGCTGCTGACGCTTCAAGCTGAGCGGCGGCGGAGAGACAGACCCCTGCAAGCA 1773
|||
Db 1080 GCGGCGCTGCTGACGCTTCAAGCTGAGCGGCGGCGGAGAGACAGACCCCTGCAAGCA 1139
QY 1774 GCGGCGCTGCTGAGCTGAGCATGATGAGAGGACGACAGCTGCGCGGATTTATGTCCTCC 1833
|||
Db 1140 GCGGCGCTTCTGAGCTGACATGATGAGAGGACGACAGCTGCGCGGATTTATGTCCTCC 1199
QY 1834 AGAGATCTTCCATTTCCACACAGCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1893
|||

```

DB 1200 AGAGATCTTCATTTCCACACGCGCTCGGATGTGCGGCTCTTACGGCATGATCTACAGCC 1259
OY 1894 CCACGCTTGACACGACGAGGAAGAGCACCACCGCTCTCTTTGATATGAGGCCCA 1953
DB 1260 CCACGCTTGACACGACGAGGAAGAGCACCACCGCTCTCTTTGATATGAGGCCCA 1319
OY 1954 GGTGACGCTGCTGATTAATCTCTTCAAAAGGCATCAAGTACTTGGCTCAACACTGGC 2013
DB 1320 GGTGACGCTGCTGATTAATCTCTTCAAAAGGCATCAAGTACTTGGCTCAACACTGGC 1379
OY 2014 CTCCCTGGGCTACGCGGTGTTGATGACGCGAGGGGCTCTCTGACGAGGGCTTCG 2073
DB 1380 CTCCCTGGGCTACGCGGTGTTGATGACGCGAGGGGCTCTCTGACGAGGGCTTCG 1439
OY 2074 GTTCGAGAGGGGCTCTCAAAACCAATGGGCGCAGGTGAGATGAGAGCAGGTGAGGG 2133
DB 1440 GTTCGAGAGGGGCTCTCAAAACCAATGGGCGCAGGTGAGATGAGAGCAGGTGAGGG 1499
OY 2134 CCTGCACTTCTGCTGCGGAGAGATGAGCTTCACTGACCTGAGCCGAGTTCATCATGG 2193
DB 1500 CCTGCACTTCTGCTGCGGAGAGATGAGCTTCACTGACCTGAGCCGAGTTCATCATGG 1559
OY 2194 CTGCTGCTACGCGGGGCTCTCTCTGCTCATGAGGGCTAATCCACAAGCCCGAGTTCATCA 2253
DB 1560 CTGCTGCTACGCGGGGCTCTCTCTGCTCATGAGGGCTAATCCACAAGCCCGAGTTCATCA 1619
OY 2254 GGTGCGCATGCGCGGCTGCGGCTGCAACCGCTGATGAGCGCTTACAGACAGAGTACACTGA 2313
DB 1620 GGTGCGCATGCGCGGCTGCGGCTGCAACCGCTGATGAGCGCTTACAGACAGAGTACACTGA 1679
OY 2314 GCGCTACATGAGCTGCTCTGAGAACACACAGAGCGCTATGAGCGGGGTTCCGTGGCCCT 2373
DB 1680 GCGCTACATGAGCTGCTCTGAGAACACACAGAGCGCTATGAGCGGGGTTCCGTGGCCCT 1739
OY 2374 GCACGTGGAAGACCTGCTCCCAATGAGCCCAACCGCTGCTAATCTCCACGCGCTTCCCGGA 2433
DB 1740 GCACGTGGAAGACCTGCTCCCAATGAGCCCAACCGCTGCTAATCTCCACGCGCTTCCCGGA 1799
OY 2434 CGAAACGCTGCACTTTTCCACACAAACTTCTGCTCTCCCACTGATCCGACGAGGAA 2493
DB 1800 CGAAACGCTGCACTTTTCCACACAAACTTCTGCTCTCCCACTGATCCGACGAGGAA 1859
OY 2494 ACCTTACAGCTCCAGATCTACCCC 2518
DB 1860 ACCTTACAGCTCCAGATCTACCCC 1884

RESULT 4
AK054656
LOCUS AK054656 2261 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ30094 fls, clone BNGH41000034, weakly similar
to DIPEPTIDYL PEPTIDASE IV (EC 3.4.14.5).
ACCESSION AK054656
VERSION AK054656.1 GI:16549242
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens neuroglioma cell_line:H4 CDNA to mRNA,
clone_11b:BNGH41 clone:BNGH41000034.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS Nishi,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A.,
Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K.,
Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y.,
Sekine,M., Kiyuchi,H., Kanda,K., Wagatsuna,M., Murakawa,K.,
Kawahori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A.,
Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K.
and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2261)
AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.

TITLE Direct Submission
JOURNAL Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'-6 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES
source Location/Qualifiers
1..2261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="BNGH41000034"
/cell_line="H4"
/cell_type="neuroglioma"
/clone_11b="BNGH41"
/note="cloning vector: pME185FL3"
539..1477
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB70784.1"
/db_xref="GI:16549243"
/translation="MSQNFDMFVSHSYSVSPCVHYVKLSGDDDLHKQPRFNASM
MEASCPDYVPPEIFPHFHTSDVRLYGYVIRPMLDPEKRIPTLVFYGVQVQLVN
NSFKIKYLRMTLASIGAYVYIDRGSCGSLFEGALKRMQVLEFDEVDLELOF
VAEKYGFIDLSVVAIHGMSYSGFSLMGLIHNPQVFKVAIACAPVTYMAAYDTGTER
YMDVPENNHGYEAGSVALLHVERLDEPRRLILILGFLDENHFHTNLVSQILIRAG
KPYQLOIYFPEHNSIRCPESGEVETTLHLFLOEYL"

BASE COUNT 470 a 704 c 640 g 447 t
ORIGIN

Query Match 56.5%; Score 1477.4; DB 9; Length 2261;
Best Local Similarity 99.9%; Pred. No. 8e-242;
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1126 CCAGAGTGAGCTCAGCTGCTCTCTCCCGCGGCGCTTTCATCCGACAGAGAA 1185
DB 1 CCAGAGTGAGCTCAGCTGCTCTCTCTCCCGCGGCGCTTTCATCCGACAGAGAA 60
OY 1186 TGAGAGACAGCGGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1245
DB 61 TGAGAGACAGCGGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
OY 1246 GTACGAGGAGGTACCAAGCTGTGATCAATGTTCATGACATTTCTATCCCTCCCA 1305
DB 121 GTACGAGGAGGTACCAAGCTGTGATCAATGTTCATGACATTTCTATCCCTCCCA 180
OY 1306 ATGAGAGGAGAGAGAGAGCTGCTTTCGCGCAATGATGCAAGACGGCTTCTG 1365
DB 181 ATGAGAGGAGAGAGAGAGCTGCTTTCGCGCAATGATGCAAGACGGCTTCTG 240
OY 1366 CCAATTTGTACAAAGTACCGCGCTTTTAAATCCAGGGCTACGATTGAGAGCCCTT 1425
DB 241 CCAATTTGTACAAAGTACCGCGCTTTTAAATCCAGGGCTACGATTGAGAGCCCTT 300
OY 1426 CACCGCGGGAAGATGAATTAAGTCCCATTAAGGAAGATGCTCTGACACGCG 1485
DB 301 CACCGCGGGAAGATGAATTAAGTCCCATTAAGGAAGATGCTCTGACACGCG 360
OY 1486 TGAATGGAGGTTTGGCGAGGACGCTCAAGATCTGGGTCAATGAGAGACCAAGCT 1545
DB 361 TGAATGGAGGTTTGGCGAGGACGCTCAAGATCTGGGTCAATGAGAGACCAAGCT 420
OY 1546 GGTGTAATTCACAGGACCAAGGACACGCGCTGAGACACACCTTACGTGCTACGTA 1605
DB 421 GGTGTAATTCACAGGACCAAGGACACGCGCTGAGACACACCTTACGTGCTACGTA 480
OY 1606 TGAGCGGCGCGGAGATGCTACGCTCACACGCGCGGCTTCCCATAGCTGCTCAT 1665

Db	481	TGAGCGGCGCGCGGATGCTATGACGCTCAGCACACGCGGGCTCTCCATPAGTGTGCAT	540
Qy	1666	GAGCCAGAACTTCGACATGTTGCTAGCCACTACAGCAGCGTGAGCAGCGGCGCTGGCT	1725
Db	541	GAGCGAAGAACTTCGACATGTTGCTAGCGCACTACAGCAGCGTGAGCAGCGGCGCTGGCT	600
Qy	1726	GCACGTCTACAAAGCTGAGCAGGCGGCCGAGCAGAGACCCCGACAAAGCAGCAGCCCGTTCG	1785
Db	601	GCACGTCTACAAAGCTGAGCAGGCGGCCGAGCAGAGACCCCGACAAAGCAGCAGCCCGTTCG	660
Qy	1786	GCGTAGCATGTATGAGGAGCGCAGCGCTGCGCCCGGATTATGTTCTCCAGAGATTTCCA	1845
Db	661	GCGTAGCATGTATGAGGAGCGCAGCGCTGCGCCCGGATTATGTTCTCCAGAGATTTCCA	720
Qy	1846	TTTTCCACACGCGCTCGATGTGCGGCTCTACGGCATGTCTTACAAGCCCCACGCTTGCA	1905
Db	721	TTTTCCACACGCGCTCGATGTGCGGCTCTACGGCATGTCTTACAAGCCCCACGCTTGCA	780
Qy	1906	GCGAGGAAAGAACGCCACGCTCCTCTTTGATATGAGGCCCCGAGGCGTACACTGGT	1965
Db	781	GCGAGGAAAGAACGCCACGCTCCTCTTTGATATGAGGCCCCGAGGCGTACACTGGT	840
Qy	1966	GAATTACTCTCTTCAAAAGGCATCAAGTACTTGGCGCTCAACACACTGAGCCTCGGGCTA	2025
Db	841	GAATTACTCTCTTCAAAAGGCATCAAGTACTTGGCGCTCAACACACTGAGCCTCGGGCTA	900
Qy	2026	CGCCGTGTTGATGATTACGGCAGGCGCTCCTGTACGAGAGGCTTCGGTTGGAAGGGC	2085
Db	901	CGCCGTGTTGATGATTACGGCAGGCGCTCCTGTACGAGAGGCTTCGGTTGGAAGGGC	960
Qy	2086	CCTGAAAAACCAATAGGGGCGAGGTGAGATGAGAGACCAAGGTGAGGGGCTGCAATTGGT	2145
Db	961	CCTGAAAAACCAATAGGGGCGAGGTGAGATGAGAGACCAAGGTGAGGGGCTGCAATTGGT	1020
Qy	2146	GCGCGAAGATATGAGCTTCATCGACCTGAGCGAATTCGCATCGATGCGTGGTCTTACGG	2205
Db	1021	GCGCGAAGATATGAGCTTCATCGACCTGAGCGAATTCGCATCGATGCGTGGTCTTACGG	1080
Qy	2206	GGGCTTCCTCTCGCTCATATGGGGCTAATCACAGGCCCAAGGTGTCAAGGTGCGCATGCG	2265
Db	1081	GGGCTTCCTCTCGCTCATATGGGGCTAATCACAGGCCCAAGGTGTCAAGGTGCGCATGCG	1140
Qy	2266	GGGTGCCCCGGGTCACCGTCTGGATGGCTTACAGACAGAGGTTCACTGAGCGCTACATGGA	2325
Db	1141	GGGTGCCCCGGGTCACCGTCTGGATGGCTTACAGACAGAGGTTCACTGAGCGCTACATGGA	1200
Qy	2326	CGTCCCTGAGAACACACAGCAGCGCTATGAGCGGGTTCCGTGGCCCTGCACGTGAGANA	2385
Db	1201	CGTCCCTGAGAACACACAGCAGCGCTATGAGCGGGTTCCGTGGCCCTGCACGTGAGANA	1260
Qy	2386	GCTGCGCAATGAGCCCAACCGCTTCTCTTATCTCTCACAGGCTCTCTGAGCAAAACGTGCA	2445
Db	1261	GCTGCGCAATGAGCCCAACCGCTTCTCTTATCTCTCACAGGCTCTCTGAGCAAAACGTGCA	1320
Qy	2446	CTTTTTCACACAACCTCTCTGTTCTCCCACTGATCGAGAGGGAAACCTTTACAGCT	2505
Db	1321	CTTTTTCACACAACCTCTCTGTTCTCCCACTGATCGAGAGGGAAACCTTTACAGCT	1380
Qy	2506	CCAGATCTACCCCAACGAGAGACACAGTATTCGTCGCCCGAGTGGGGGAGACACTATGA	2565
Db	1381	CCAGATCTACCCCAACGAGAGACACAGTATTCGTCGCCCGAGTGGGGGAGACACTATGA	1440
Qy	2566	AGTCAGTTGCTGCACCTTTCTACAGAGAACTCTTGAGC 2604	
Db	1441	AGTCAGTTGCTGCACCTTTCTACAGAGAACTCTTGAGC 1479	

RESULT 5	BC000970	BC000970	2546 bp	mRNA	linear	PRI 12-JUL-2001
LOCUS	BC000970					
DEFINITION	Homo sapiens, clone IMAGE:3447394, mRNA, partial cds.					
ACCESSION	BC000970					
VERSION	BC000970.1		GI:12654298			

KEYWORDS	
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 2546)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (16-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAK Plate: 3 Row: 1 Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES	
source	Location/Qualifiers
	1..2546
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:3447394"
	/tissue_type="Placenta, choriocarcinoma"
	/clone_1lb="NH_MGC_10"
	/lab_host="DH10B"
	/note="Vector: pCMV-SPORT6"
	<1..1320
CDS	/codon_start=1
	/product="unknown (protein for IMAGE:3447394)"
	/protein_id="AAH00970.1"
	/db_xref="GI:12654299"
	/translation="DIFYPPQSEGEDELFLRANECKTGFCHLYKTAVALKSQGYDM SEVSPDEDEKCIKEEIALTSGEWMILRHGSK IWNDETKLYVQGRKDPLEHH LYVSYAAGEIVRLTTPGFSHSCSMONMDLPHGSHSVSTPCYVAVYKSGDDP LHKOPRWASMEARASCPDPVYPPETIHFHTRSYRLIKGATYKFAHLOPKRIITVLF VYGGPQVQLVNNSEFKIGIKYLRKLTNLASLGAVVVIDRGSQGRLEKREGALIKMGQV EIEQVQVGLQFAVEKYGFIKLSRVAIHGWSYGGFLSLMGILHKQVYKFAIAGAPTV IMAAIDGYTERYMDVYPENNOHGYEAGSVLAHVEKLPENPRLLILHFELEDNVAFHT NFLVSQLIRAGKPVQLOIYPRNERSINCPESGEHYETLHFLQEYL"
BASE COUNT	556 a 795 c 714 g 481 t
ORIGIN	
Query Match	50.5%; Score 1322; DB 9; length 2546;
Best Local Similarity	100.0%; Pred. No. 2.2e-215;
Matches 1322; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Yy	1283 GACATCTTCTATCCCTTCCCTCCCAATCAGAGGAGAGAGAGACTCTGCTTCTCCGGCGC 1342
Dd	1 GACATCTTCTATCCCTTCCCTCCCAATCAGAGGAGAGAGAGACTCTGCTTCTCCGGCGC 60
Oy	1343 AATGATGCAAGACCGCGCTTGCCTATTGTACAAAGTCACCGCGCTTTAAATCCAG 1402
Dd	61 AATGATGCAAGACCGCGCTTGCCTATTGTACAAAGTCACCGCGCTTTAAATCCAG 120
Oy	1403 GGCTACGATTGGAGTGAACCCCTTAGCCCCGGGGGAGATGAAATTTAAAGTCCCATTAAG 1462
Dd	121 GGCTACGATTGGAGTGAACCCCTTAGCCCCGGGGGAGATGAAATTTAAAGTCCCATTAAG 180

QY 1463 GAAGAGATTGCTCTGACACGCGGTGAATGGAGTTTGGCGAGGCAAGGCTCCAAAGATC 1522
|||||
Db 181 GAAGAGATTGCTCTGACACGCGGTGAATGGAGTTTGGCGAGGCAAGGCTCCAAAGATC 240
QY 1523 TGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCCAAAGACAGCGCGCTGGAG 1582
|||||
Db 241 TGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCCAAAGACAGCGCGCTGGAG 300
QY 1583 CACACACTCTACCTGTGACGCTATGAGGGCGCGGAGATCGTACGCGCTCACACGCGC 1642
|||||
Db 301 CACACACTCTACCTGTGACGCTATGAGGGCGCGGAGATCGTACGCGCTCACACGCGC 360
QY 1643 GCGTCTCCCAATGACTGTCTCCATGAGCCAGCACTTGTGCTGACGCACTACAGC 1702
|||||
Db 361 GCGTCTCCCAATGACTGTCTCCATGAGCCAGCACTTGTGCTGACGCACTACAGC 420
QY 1703 AGGTGAGCAGCGCGCGCTGCGCTGACAGCTTACAAGCTGAGCGCGCGCGGAGACGCC 1762
|||||
Db 421 AGGTGAGCAGCGCGCGCTGCGCTGACAGCTTACAAGCTGAGCGCGCGCGGAGACGCC 480
QY 1763 CTGCACAGCAGCGCGCGCTTCTGGGCTAGCATGATGAGGAGCAGCAGCTGCGCGGAT 1822
|||||
Db 481 CTGCACAGCAGCGCGCGCTTCTGGGCTAGCATGATGAGGAGCAGCAGCTGCGCGGAT 540
QY 1823 TATGTTCTCCAGAGATCTTTCATTTCCACAGCGCTGGATGCGGCTCTACGGCATG 1882
|||||
Db 541 TATGTTCTCCAGAGATCTTTCATTTCCACAGCGCTGGATGCGGCTCTACGGCATG 600
QY 1883 ATCTAACAAGCCCGACGCTTTCGACGCGGAGGAAAGACAGCCGCTCTTGTATAT 1942
|||||
Db 601 ATCTAACAAGCCCGACGCTTTCGACGCGGAGGAAAGACAGCCGCTCTTGTATAT 660
QY 1943 GGAGGCGCGCGAGGTGAGCTGTGAATTAATCTCTCAAGAGCATCAAGTCTTGGCGCTC 2002
|||||
Db 661 GGAGGCGCGCGAGGTGAGCTGTGAATTAATCTCTCAAGAGCATCAAGTCTTGGCGCTC 720
QY 2003 AACACATGGGCTCCCTGGGCTACGCGCTGTGTATGATGAGGAGGCGCTCTGTACG 2062
|||||
Db 721 AACACATGGGCTCCCTGGGCTACGCGCTGTGTATGATGAGGAGGCGCTCTGTACG 780
QY 2063 CGAGGCGCTTCTGTCGAAGGGGCGCTGAAAAACCAATGGGCGAGGTGAGATCGAGAC 2122
|||||
Db 781 CGAGGCGCTTCTGTCGAAGGGGCGCTGAAAAACCAATGGGCGAGGTGAGATCGAGAC 840
QY 2123 CAGGTGAGGCGCTGAGTCTGTGGCGGAGAAATGAGCTTCAATCGACCTGAGCCGAGT 2182
|||||
Db 841 CAGGTGAGGCGCTGAGTCTGTGGCGGAGAAATGAGCTTCAATCGACCTGAGCCGAGT 900
QY 2183 GCCATCATGCTGTGTCTTACGCGGCGCTTCTCTGCTCATGGGCGTAAATCCAAAGCCC 2242
|||||
Db 901 GCCATCATGCTGTGTCTTACGCGGCGCTTCTCTGCTCATGGGCGTAAATCCAAAGCCC 960
QY 2243 CAGGTTCAGGTGGGCTACGCGGCGCTGCGCGGCTACCGCTGTGATGGCTTACGACACA 2302
|||||
Db 961 CAGGTTCAGGTGGGCTACGCGGCGCTGCGCGGCTACCGCTGTGATGGGCTTACGACACA 1020
QY 2303 GGGTACACTGAGGCTACATGAGCTGCCGAGAACACAGCAGGCTATGAGGGGGGT 2362
|||||
Db 1021 GGGTACACTGAGGCTACATGAGCTGCCGAGAACACAGCAGGCTATGAGGGGGGT 1080
QY 2363 TCCGTGGCGCTGACGTGAGAGCTGCCCAATGAGCCCAACCGCTTGTATCTCCAC 2422
|||||
Db 1081 TCCGTGGCGCTGACGTGAGAGCTGCCCAATGAGCCCAACCGCTTGTATCTCCAC 1140
QY 2423 GCGTCTCTGAGAGAAAGCTGCTTTTTCACACAACTTCTGCTTCCCACTGATC 2482
|||||
Db 1141 GCGTCTCTGAGAGAAAGCTGCTTTTTCACACAACTTCTGCTTCCCACTGATC 1200
QY 2483 CGAGCAGGAGAACTTACCAAGCTCCAGATCTACCCCAAGAGAGACAGATATTCGCTGC 2542
|||||
Db 1201 CGAGCAGGAGAACTTACCAAGCTCCAGATCTACCCCAAGAGAGACAGATATTCGCTGC 1260

QY 2543 CCCGAGTCGGGCGAGCAGCTATGAAATGATCAGCTGCTGCACCTTCTTACAGGAATACCTCTGA 2602
|||||
Db 1261 CCCGAGTCGGGCGAGCAGCTATGAAATGATCAGCTGCTGCACCTTCTTACAGGAATACCTCTGA 1320
QY 2603 GC 2604
Db 1321 GC 1322
RESULT 6
AX354795
LOCUS AX354795 2649 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 3 from Patent WO01/9473.
ACCESSION AX354795
VERSION AX354795.1 GI:18619528
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Meyers, R.A. and Williamson, M.
TITLE 21953, a human prollyl oligopeptidase family member and uses thereof
JOURNAL Patent: WO 01/9473-A 3 25-OCT-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1..2649
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 803 a 514 c 586 g 746 t
ORIGIN
Query Match 33.0%; Score 862.6; DB 6; Length 2649;
Best Local Similarity 59.4%; Pred. No. 3,7e-137;
Matches 1500; Conservative 0; Mismatches 1019; Indels 6; Gaps 2;
QY 80 TTCCAGGTGAGAGCAGCTGTGAGAGCGGCTCCGAGCATCATCCAGCGGACCCCAAG 139
|||||
Db 103 TTTTATGTTGAGCGGTATCTCTGAGTTCAGCTTAAAAAGCTGCTCCGCTACAGAAAA 162
QY 140 TACTCGGCGCTCATTTGCAACAGGGCGCCAGCAGCTTCCAGTTTGCGAAGAGGAT 199
|||||
Db 163 TATCATGCTGCTACATGATGCTTAAGGACACATGATTTCTATGTTTGGAAGAGAGAT 222
QY 200 GAGTCTGGGCGCCACTCCACCGCTCTACTACTCTGGAATGCAATGAGCAGCGAGAG 259
|||||
Db 223 CCAGATGAGACCTATTCAGCAAGATCTATTAACCTTGCCATGTCTGTGAGAACAGAGAA 282
QY 260 AACTCCCTCTCTACTCTGATTTCCAGAGAGGTCCGGAAGAGGCTGTGCTCTG 319
|||||
Db 283 AATACACTGTTTATTCGAAATTCACCAAAATCAATAGACACAGCTTAATAGCTC 342
QY 320 TCTGTGAGAGAGATGCTGATTCATTTCCAGGGCGACGCCACCATGGGGCTCTCTCG 379
|||||
Db 343 TCTTGAAGGCTCTTGTGATTTCTTTTCAAGGCAACCTGAGCTATGGAATGATTTCTGA 402
QY 380 GAGGAGAGCTGCTGAGGAGCGGAAAGCGCTGAGGCTTCTGCGCATCACTCTCAAGAC 439
|||||
Db 403 GAGGAGAGCTATTAAGAGAGAAAGAAAGCATTTGGAAGCATGCGAATGTCTTACGAT 462
QY 440 TTCCACAGCAGAGAGTGCGCTCTCTTCCAGGCGCAGCAACAGCTCTTCCACGTGCG 499
|||||
Db 463 TATCACCAAGAGAGTGAACATTTCTGTTTCAAGCGGAGTGAATTTATTCACGTAAAA 522
QY 500 GAGGCGGCGCAAGAGCGCTTCAATGATGTCCTCATTAACCCGTGGAAATCAAGACCCAG 559
|||||
Db 523 GATGAGGCGCCACAGAGATTTACGCAACAACCTTTAAGGCCCAATCTATGAGGAACTAAG 582
QY 560 TGCTCAGGGCGCCGATGAGCAACAAATCTGCGCTGCGGACCGCTCTCTCTCTTC 619
|||||
Db 583 TGTGCCAACAATAGATGATGCCAAATATATGCGCTGTGATCAGACATGAGATTTGCTTT 642
QY 620 ATCAATTAACAGCACTGTGGGTGGCCAAATCAGAGACAGGCGAGGCGGCTGACC 679

Db	1717	AGTAACAGAAAGAAATCCACACTGCTGTGTCCTTTACAAAGCATCAAAAGTCAGAGATAC.	1776
QY	1760	CCCCGCACAAGACAGCCCCGCTTCTGGGCTAGCATGATGAGGACGCCAGCTGCCCCCG	1819
Db	1777	CCAACCTTCCAAACAAAGAAATTTTGGGCCACCATTTTGGATTCCAGAGTCCTCTTCT	1836
QY	1820	GATTATGTTCTCCAGAAATCTTCATTTCCACACGGCCCTGGAGTGTGGGCTCTACGGC	1879
Db	1837	GACTATCTCTCTCCGAAATTTTCTCTTTTGAAGATCTACTGATTTACATTTGATGGG	1896
QY	1880	ATGATCTCAAGACCCCGACCTTTCGAGCCAGGAGAAAGACACCCACCCGCTCTTTGTA	1939
Db	1897	ATGCTCTCAAGACCTCATGATCTACAGCCTGGAAGAAATATCTCATCTGCTGTTCATTA	1956
QY	1940	TATGAGAGCCCCCAGAGTGCAGCTGTGTGAATATCTCTTCAAGGCATCAAGTACTTGGCG	1999
Db	1957	TATGTGTGCTCTCAGATGCTGATGTGTGAATATCGGTTTAAAGAGATCAAGTATTTCCG	2016
QY	2000	CTCAACACACTGAGCTCCCTGGGCTACGCCCGGTGTGTGATTTAGAGGACAGGGGCTCTGT	2059
Db	2017	TTGAATTAACCTTAAGCTCTCTAGGTTATGTGTGTGAAGATAGACAAACAGGGATCCGT	2076
QY	2060	CAGCAGAGGCTTCCGTTTGAAGGGGCCCTGAAAAACCAATAGGGCAGGTGAGATCGAG	2119
Db	2077	CACCGAGGGCTTAAATTTGMAAGGCCCTTTAAATTTAAATATGGTCAATATGAATTTGAC	2136
QY	2120	GACCAAGTGAAGGGGCTGCAGTTCTGTGGCCGAGAAATATGCTTCATGACCTGAGCCGA	2179
Db	2137	GATCAAGTGAAGGAGACTCCAAATATCTAGCTTCTGATATGATTTGATTTAGATATGAT	2196
QY	2180	GTTGCCATTCATGCTGCTGCTCTACGGGGGCTTCTCTCTGCTCATGTAGGGGCTAATCCAG	2239
Db	2197	GTTGGCATTCACAGGCTGTGCTCATGTGAGGATATCCCTCTCCATGATGCAATTAATGACAG	2256
QY	2240	CCCCAGGTTCAAGGTGGCCATTCGGGGGTCCCGGGTCAACCGTGCATGGGCTACAGC	2299
Db	2257	TCAGATATCTTAGGGTGTCTATTTCTGGGGCCCAAGTCACTCTGTGGATCTTCTATGAT	2316
QY	2300	ACAGGATCACTGACGCGTACATGAGACGTCCTTGAAACACACACAGCGCTATGAGGCG	2359
Db	2317	ACAGGATCAACAGGAACGTTATATGTGGTCACTCTGACACAGAAATGAACAGGGCTATTACTTA	2376
QY	2360	GCTTCCGTGGCCCTGCACGTGAGAGAGCTGCCCAATGAGCCCAACCGCTTGTCTCTC	2419
Db	2377	GGATCTGTGGCCATCAAGCAGAAAAAGTTCCCTCTGAACCAAAATCGTTTACTGCTCTTA	2436
QY	2420	CACGGCTCCCTGAGAGAAAGCTGCACCTTTTCCACACAAACTTCCCTGCTCCCAACG	2479
Db	2437	CATGGTTCTCTGATGAGATGTCTCATTTTTCACATATCCAGTATTTTACTAGATTTTTTA	2496
QY	2480	ATCCAGAGCAGGAAACCTTACCAAGCTCTCCAGATCTACCCCAACGAGAGACACAGTATTCG	2539
Db	2497	GTAAGGGCTCGAAGAACATATGATTTTACAGATCTATTCCTCAGAGAGACACAGCATTAAGA	2556
QY	2540	TGCCCGAGTCGGGCGAGCATATGAAGTACAGTTGCTGCACCTTCTACAGGAATACCTC	2599
Db	2557	GTTCCCTGAATCGGAGAAACATTAATGAACATGCACTCTTTTCCACTACCTTCAMAAAAACCTT	2616
QY	2600	TGAGC 2604	
Db	2617	GGATC 2621	
RESULT 7			
LOCUS	AX342633	3106 bp	DNA linear PAT 12-JAN-2002
DEFINITION	Sequence 30 from Patent WO0198468.		
ACCESSION	AX342633		
VERSION	AX342633.1	GI:18152030	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

QY	860	GAGTGGAGGTCATTCACGTCGCCCTCTCTGGCTTGAAGAAGAGACGAGACTCGAT	919
Db	1108	GAGGCGAAATTAATTCATGTTACATCCCTATTTGTGAACAAGAGGGCCATATTCATTC	1167
QY	920	CGGTACCCACGACAGGACAGCAAGAAATCCCAATTCGCTGGAACCTGGCTAGTTCCAG	979
Db	1168	CGTTATCTTAAACAGGTACACGCAAAATCCTAAGTAGCTATTTAAGATGTCAAGAAATATG	1227
QY	980	ACTGACAGCCGAGGACAGATCGTCTGACCCAGAGAGAGAGAGCTGGTGACGCCCTTCAGC	1039
Db	1228	ATTGATCGTGAAGAGAGATCATAGATGTCATAGATAAGAACTAATTAACACTTTTGAG	1287
QY	1040	TGCGGTCCCGGAAGGTGGAGTACATCCGACGGCCGGGTGGACCCCGGATGGCAAAATAC	1099
Db	1288	ATTCTATTGAAGAGATTTGAATATATTTGCCAGAGCTGGATGAGACTCTGAGGGAATAAT	1347
QY	1100	GCTTGCGGCATGTCCTGAGACCGGCCCCAGACAGATGGCTCAGCTGCTCCTCCGCCG	1159
Db	1348	GCTTGCTCATCTCTACTAAGATCGGCTCCACAGCTGGCTGAGATGTTGATCTACACT	1407
QY	1160	GCCCTGTTCATCCGACACAGAAATGAGAGACAGCGCTAGCCTCTGCCAGAGCTTC	1219
Db	1408	GAATATTATTATCCAGTAGAAGATGATGTTATGAAAGGACAGACACTATTGAGCACTG	1467
QY	1220	CCAGGAATGTCCACCGCTATGTGTGTACAGAGAGGTACACACGTCTGGATCAATGTT	1279
Db	1468	CCTGATTTCTGAGAGCCACATATATCTATGAAGAAACAACAGACATCTGATTAATATC	1527
QY	1280	CATGACATCTTATTCCTTCCTCCCAATCAGAGGAGAGAGACGAGCTGCTTCTCCG	1339
Db	1528	CATGACATCTTTCATGTTTTTCCCAAGACAGGAAG---GAATATGAGATTTATTTT	1584
QY	1340	GCCATGAATCAGAGACCGGCTCTGTCATTTTACAAAGTCAACGGCTTTAATAATCC	1399
Db	1585	GCTCTGATTCACAAACAGGTTTCGTCATTTATACAAATATATCTATTTTAAAGGA	1644
QY	1400	CAGGCTACGATTGGAGTGAGCCCTTCAGCCCCGGGGAAGATGAATTTAAGTCCCCAT	1459
Db	1645	AGCAAAATTAACAGATCAGTGGGGGCTCCTGCTCAAGATGATTTCAAGTGTCTATC	1704
QY	1460	AAGGAAGATTTGCTCTGACCACGGTGAAATGGAGAGTTTGGCCAGAGACGGCTCAAG	1519
Db	1705	AAAGGAGGATAGCAATTTACCAGTGGTGAATGGGAAGTTCTTGGCCGGATGAGATCTAT	1764
QY	1520	ATCTGGGCAATGAGAGACCAAGCTGTGTACTTCACAGGACACAAAGACAGCGCGTG	1579
Db	1765	ATCCAAAGTTGATGAAGTCAAGAAAGCTGTGATATTTTGAAGGCAACAAAGACTCCCCTTA	1824
QY	1580	GAGCACCACTCTACGCTGTCAGCTATGAGGGCGGCGGAGATGTGTACGCTTACACAG	1639
Db	1825	GAGCATCACCTGTACGTATGTCAGTTAGTAAATCTCTGGAGAGGTGCAAGGCTGTACTAC	1884
QY	1640	CCCGGCTTCTCCATTAAGCTGTCATGAGCAGACATTTGACATGTGTGTCAGCCATAC	1699
Db	1885	CGTGGCTACTACATTTCTTGCTGATCAGTCAACACTGTGACTCTTTTATAGTAAGTAT	1944
QY	1700	AGCAGCGTAGACAGCGCCCTGCGTGCACTACAGCTGAGCGGCCCGCAGCAGAC	1759
Db	1945	AGTAACCGAAGATTCACACTGTGTGTCCCTTTACAACTATCAAGTCTTGAAGATAG	2004
QY	1760	CCCCGCAAGAGCGCCGCTTCTGGGCTAGATATAGGAGGACGACAGCTGCCCCCG	1819
Db	2005	CCAACTTGCAAAACAAAGAAATTTTGGGCCACCATTTTGGATTTACAGAGCTCTTCTCT	2064
QY	1820	GATTATGTTCTCCAGAGATCTTCATTTTCCACAGCCGCTCGGATGTGGGCTCTACGCG	1879
Db	2065	GACTATACTCTCCAGAAATTTCTGTTTTGAAGTACTACTGTGAATTTACATTTGATAGG	2124
QY	1880	ATGATCTACAGCCCAAGCCCTTGACAGCCAGGAGAGACACCCACAGCTCTCTTTGTA	1939
Db	2125	ATGCTCTACAGGCTCTATGATCTACAGCCCTGGAAAGAAATATCTACTGTGCTTTCTA	2184

QY	1940	TATGGAGGCCCCCAGGTGACGTGTGTAATAACTCTTCAAAAGCATCAACTACTGTGGG	1939
Db	2185	TATGGTGTGCTTCAGAGGTGCAGTGGTGAATATATCGGTTTTAAAGGAGTCAACTATTTTCCGC	2244
QY	2000	CTCAACACACTGAGCCCTCCCTGGCTACGCCCGGTGTGTGATTGACGCGAGGGCCCTGT	2059
Db	2245	TTTGATTAACCCCTACGCTCTCTAGTTATATGTGTGTAGTAGACACACAGGGGATCCGT	2304
QY	2060	CAGCAGGAGGCTTCGGTTCGAAAGGGCCCTGAAAAACAATATGGCCAGGTGAGATCGAG	2119
Db	2305	CACCCAGGGGCTTAAATTTTGAAGGCCCTTTTAAATTTAAAAATGGGCTCAAAATGAAATTTGAC	2364
QY	2120	GACCAAGTGGAGGGGCTCGAGATTCGTGTGGCCGGAAGATTTGGCTTCATCGACTGAGCCGA	2179
Db	2365	GATCAAGGTGGAAGAGCTCCAATATCTACTCTCTGATATGATTTTCATTGACTTAGATCGT	2424
QY	2180	GTTGGCATTCATAGCTGGTGCCTACGGGGGCTTCCTCTCGCTCANTGGGGCTAATCCACAAG	2239
Db	2425	GTTGGCATTCACAGCGCTGGTCCATGTGAGGATACCTCTCCCTGATGCACTTAATCCAGAGG	2484
QY	2240	CCCCAGGTTCGAAGTGGCCATTCGCGGGGTGCCCCGCTACCGTCTGGATGGCCATGAC	2299
Db	2485	TCAGTATCTTCAGGGTGTGCTATCTGTGGGCCCGCCAGTCACTCTGTGGATCTTCTATGAT	2544
QY	2300	ACAGGATCACTGACGAGCGCTACATGAGACGCTCCCTGAAACAACACAGCGCTATGAGCG	2359
Db	2545	ACAGGATCAACAGGAACGTTATATGGGTCAACCCTGCACAGAATGAACAGGGCTATTACTTA	2604
QY	2360	GGTTCCTCGTGGCCCTACAGTGTGAGAAAGCTGCCCAATGAGCCCAACCGCTTGTACTCTC	2419
Db	2605	GGATCTGTGGCCATCAACAGCAAAAAAGTTTCCCTCTGAAACAAATCGTTTACTGCTCTTA	2664
QY	2420	CACGGCTTCCTGAGAGAAACGTGCACCTTTTCCACACAAACTTCCTGTCGCCAACGTG	2479
Db	2665	CATGGTTTTCCTGTGATGAGAAATGTCCATTTTGTGCACATACCAATATATTACTAGTTTTTA	2724
QY	2480	ATTCGACGACGGAACCTTTACCACTCCAGATCTACCCCAAGAGACACACAGTATTCGC	2539
Db	2725	GTAGAGGGCTGGAACCACTATGATTTTACAGATCTATCTCCAGAGAGACACAGCTTAAGA	2784
QY	2540	TTCGCCGATGGGGGAGCAGCTATGACATGACAGTGTGCTCACTTTCACAGAAATACCTC	2599
Db	2785	GTTCTCTGATCGGAGAACATTATTAAGTGCATCTTTTGCACCTTACCTTCAAGAAACCTT	2844
QY	2600	TTGAGC 2604	
Db	2845	GGATC 2849	

RESULT 9					
AF221634					
LOCUS	AF221634	3127 bp	mRNA	linear	PRI 05-NOV-2000
DEFINITION	Homo sapiens dipeptidyl	peptidase 8 (DPP8)	mRNA, complete cds.		
ACCESSION	AF221634				
VERSION	AF221634.1	GI:11095187			
KEYWORDS					
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 3127)				
AUTHORS	Abbott,C.A., Yu,D.M., Woollett,E., Sutherland,G.R., McCaughan,G.W.				
	and Gorrell,M.D.				
TITLE	Cloning, expression and chromosomal localization of a novel human				
JOURNAL	dipeptidyl peptidase (DPP) IV homolog, DPP8				
MEDLINE	Eur. J. Biochem. 267 (20), 6140-6150 (2000)				
PUBMED	11012666				
REFERENCE	2 (bases 1 to 3127)				
AUTHORS	Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-JUN-2000) A.W. Morrow Gastroenterology and Liver				
	Centre, Centenary Institute of Cell Biology and Cancer Medicine,				

FEATURES		Locked Bag No 6, Newtown, Sydney, NSW 2042, Australia	
source	Location/Qualifiers		
	1. .3127		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="15"		
	/map="15q22"		
gene	1. .3127		
	/gene="Dppg"		
	214. .2862		
CDS	/note="serine protease; member of prolyl oligopeptidase S9b family belonging to enzyme clan SC; alternatively spliced forms of this mRNA are found in AF221635, AF221636 and AF221637"		
	/codon_start=1		
	/product="dipeptidyl peptidase 8"		
	/protein_id="AACG9766.1"		
	/db_xref="GI:11095186"		
	/translation="MAAAMETEQLGEVIEFTADCEENIESODRKLPEPFYVERYSWSQ LKLLADTRKRYHGMYMKAAPDFEMFVRKNPDGPHSRIRIYLASGNRENTLFPSEI PKTINRAAVILMSKPLDLFDATIDYGAGSREEELRRERIKITGVIAISDYOGSGG TELQAGSGRIHYVKDDGGPGFTQOLPRLNVEETSCPNLRMPDKIPADPMIAIHSM DIAMSNIVTTERRLTYVHNFLAMEDBASAGVATRYLOEPRPYRGYMKCPRAETTP PSGSKILRILEEDESDEVELIHVTSPLLETRRADSFRTPTGTANRVTFKSELIM DAERI.IDIVDELKLOPELIFEGEVYIARAGMPBEKANSLILDRSORLQIVLIS PELPIPEDVDIMERORLIESVPDSVTPLIYEETTDIINHLDFHPORHEEFEE IFAEEKCKFHFKLYKITSIKESKYRSSGLPASPDPCKIKRELAITSGMEVLGR HGSNIQDEVNRILVFEFGTKDSPLEHLIYYVSVNPEVPLMRGYSHSOICSONHD FFIKRYSNOKNPHCVLSLKLSSPPDDPKTKERWATTLDSAGLPDTPTPEITSFBS TVVLTIGMLKRPDLPOPKKKYPTLVLFYGGPOVALNNRKGVKFRPLNLTLASLGIV VVVIDNRSCHRGKLFEGAFKKMGGOILEIDGVOGLQYLASRYDFIDLDRNLNGSWY AKSYLMALMQRSDIFRYALIGADVTLMIEDYTERTERYMHGPQNBOGYGISVAQM AEKPSEPNNRLMLHGFLEDENVHFAHSIILSFVLRAGKPYDLQITYPERNISIVPES GEHELHLHLHLOENLIGSRILAALVI"		
BASE COUNT	943 a 637 c 706 g 841 t		
ORIGIN			
Query Match	32.9%; score 861; DB 9; Length 3127;		
Best Local Similarity	59.4%; Pred. No. 6,7e-137;		
Matches 1499; conservative	0; Mismatches 1020; Indels 6; Gaps 2;		
OY	80 TTCCAGGTGCAGAAACACTCGTGGAAGGGCTCCGGAGCATCATTCCAGGCCAGCCGCAAG 139		
Dd	316 TTTTATGTTGAGCGGTATTCCTGAGTAGCACTTAATAAGCTGCTGGCATACAGAAAA 375		
OY	140 TACTCGGGCTCATTTGTCAAAGAAGCGCCGCCAGCACTTCCAGTTTGTGCAAGAACGGAT 199		
Dd	376 TATCATGGCTACATATGCTAGGCTAGGACCACCACTGATTTCTATGTTGTGAAGGGATGAT 435		
OY	200 GAGTGTGGGGCCCCATCCACCACCGCTTACTACTCGTGGGAATCCGATATGAGGACCGCGAG 259		
Dd	436 CCAGATGGACCTCATTCACAGACAAGAAATCTATTCCTTGGCATSTCTGGTGAGAACAGAGHA 495		
OY	260 AACTCCCTCTCTACTCTGAGATTCCCAAGAAAGTCCGGAAAAGAGCTCTGCTGCTCTCG 319		
Dd	496 AATACACGTGTTTATTCTGGAATTTCCCAAAATATCAATATAGAGCAGCAGTCTTAATGCTC 555		
OY	320 TCCTTGGAAAGACATGCTGATATTTCCAAGGCCACGCCACCATTGGGGTCTACTCTCGG 379		
Dd	556 TCTTGGAAAGCCCTTTTGGATCTTTTTTCAAGGCACACAGTGAATTAATGGAATGATTTCTGCA 615		
OY	380 GAGGAGGAGCTCTCAGGGAGCGGAAAGCCCTGGGGGTCCTTGGGATCATCACTTCCTACGC 439		
Dd	616 GAAGAAGAACTTATTAAGAGAAAGAAAAGCAATGGACACAGTGGGAATTTGCTTTACGAT 675		
OY	440 TTCACAGGAGAGTGGCCCTCTTCTCTTCCAGGCCAGCACAGCAGCCTTCCACTGTGCGC 499		
Dd	676 TATTCACCAAGAAAGTGAACATTTCTGTTTCAAGCCCGGATGGAATTAATACGTAATAA 735		
OY	500 GACGGCGGACAGAACGGCTTCATGGTGTCCCTTATGAACAACGCTGGAAATCAAGACCCAG 559		
Dd	736 GATGAGGGGCCCAAGGATTTACGACACACCTTTTAAGGCCCAATCTATGTGGAACCACTAGT 795		

OY	560	TGTCACGAGGCCCCGGATGAGACCCCCAAAATCTGCGCTCCGACCGCCCTCTCTCTCC	613
Db	796	TGTCCCAAACATACGGATGATGATCCAAAATTTATGCCCCGCTGATCCAGACTGGATTGCTTT	855
OY	620	ATCATATAACAGCGACCTGTGGTGGGCGCAACATTCGAGACAGCGGAGAGAGCGCGCGCTGAC	679
Db	856	ATATCATAGCAACGATATTTTGGATATCTCAACATGTAACCAAGAGAAAGAAAGAGACTCACT	915
OY	660	TTTCGCCACCAAGTTTATCCATGTCTCGTGATGACCCCCAAGTCTGCGGTTGTGGCAC	739
Db	916	TATGTGACACATGAGCTAGGCCAACATGGAAGAAAGATCCAGATCTAGCTGAGTGGCTAC	975
OY	740	TTTCGCTATACGAGAGATTCTCGACCGCTTCACTGGGTACTGGGTGGGCGCCACAGCTCC	799
Db	976	TTTGTCTCCAGAGAAATTTGATAGATATTTCTGGCTATTGGTGGTGTCCAAAAGCTGAA	1035
OY	800	TGGGAGAGTTCAGAGGGCCCTCAAGACGCTGCGAATCCTGTATGAGAAAGTCGATGATGTC	859
Db	1036	AC---AACTCCCACTGGTGGTAAATTTCTTAAGTAATCTATGAGAAATAATGATGAACT	1092
OY	860	GAGGTGAGAGCTCATTCACGTCCTCCCTCCGCGCTAGAAAGAAAGACGACCTGAT	919
Db	1093	GAGGTGAGAAATTATTCATGTTTACATCCCTATGTTGAAACAGAGAGGGCAGATTCATTC	1152
OY	920	CGGTACCCCAAGACGAGCAGACAGAAGATCCCAAGATTGCCCTTAACATGGCTAGTTCAG	979
Db	1153	CGTTATCCTAANAACAGGTACAGCAAAATCCTAAGTCACTTTTAAGATGTGCAGAAATATG	1212
OY	980	ACTGACACCCAGGGCAGAGATGCTCTGACCCAGAGAGAGAGAGCTGGTGCACCTTCAGC	1039
Db	1213	ATTATATGCTGAAAGAGAGATCATAGATGTGATGTGATTAAGAACTAATTCATCACCTTTGAG	1272
OY	1040	TCGCGTTCCCGAAGGTGGAGTACATCGCCACGAGGGCGGGGAGCCCGGGATGGCAAAATC	1099
Db	1273	ATTCTATTTTGAAGAGATTGAAATATTTTGGCAGAGCTGAGATGCACTCTGAGGGAAATAT	1333
OY	1100	GCTTGGGCCATGTTCTTGAGACGGGCCCCAGACAGATGGCTCCAGCTGCTCCTCCGCCCG	1159
Db	1333	GCTTGGTCCATCTCACTAGATGATCCCTCCAGACTCGCCTACAGATGATGTTGATCTCACCT	1392
OY	1160	GCCCTGTTACCCCGACACAGAGAAATGAGAGAGACAGGCTGACCTCTGCCAGAGCTGC	1219
Db	1393	GAATATTTATTTATCCAGTAGAAGATGATGTTATGGAAGAGGACAGACATATGATGATCATG	1452
OY	1220	CCCAAGGATGTCACCGCTATGTGTGTGTAAGAGAGAGTCAACCAAGTCTGATCAATGTT	1279
Db	1453	CCTGATCTCTGAGCGCACATAATATCTATGAAAGAACAAAGACATCTGGATTAATATC	1512
OY	1280	CATGACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGAGAGAGCTCTGCTTCTCCG	1338
Db	1513	CATGCACTTTTCACTGTTTTTCCCAAAAGTCAAGAAAG---GAATGTAGTTTATTTTT	1569
OY	1340	GCCATATGAATGACAGACCGGCTCTGCACTTTTACAAAAGTCAACCGCGTTTAAATATC	1399
Db	1570	GCCCTGATATCAAAACAGGTTTTCGTCATTTTACAAAATTTACATCTATTTTAAAGAA	1628
OY	1400	CAGGCTACAGTTGGAGTGAGCCCTTCAAGCCCCGGGGAAAGATGAATTTAAGTGCCTCAT	1459
Db	1630	AGCAAAATTTAAACATCACTAGTGGGGCTCTGCTCCAAAGTAAATTTCAAGTGCCTATC	1689
OY	1460	AAGGAAGATGTTGCTGACCAAGGGTGAATGGGAGGTTTGGCAGAGCGAGGCTCCAG	1519
Db	1690	AAAGGGGATAGCAATTTACACAGTGGGTAATGGAAAGTTCTTGGCCGGATGATCTAAT	1749
OY	1520	ATCTGGGTCAATGAGAGACCAACAGTGTGCTACTTCCAGGACACCAAGGACACGCGCTG	1579
Db	1750	ATCCAAGTTGATGAAGTCAGAGGCTGTAATTTTGAAGGACCAAAACATCCCTTTTA	1809
OY	1580	GAGCAACACCTCTACGTGTGACGTATAGAGCGCGCGCGGAGATCGTAGCCTCACACG	1639
Db	1810	GAGCATCCCTGTACGTAGTCACTAGTAAATCTGAGAGAGGTGCAAGGCTGTACTAC	1869

QY	1640	CCGGCTTTTCCCATACCTCTCCATGAGCAGAACTTCGACATGTGGTCAGCACAC	1639
Db	1870	CGTGGCTACTACATTTTCTCTGCTATCAGTCACACACTGTGACTTCTTTATATGATGAT	1929
QY	1700	AGCAGCGTGAGCACAGCCGCTCTGTCGACAGTCCTACAGCTGAGCGGCCCGCAGACAC	1759
Db	1930	AGTAACCAAGAAAGATCCACACTGTGTGTCCCTTTACAACTATCAAGTCTGAAAGATAC	1989
QY	1760	CCCTGCACAAAGACGCCCTCTCTGGCTAGCATGATGAGGCGACCGACTGCCCCCG	1819
Db	1990	CCAACTTGCAAAACAAAGAAATTTTGGGCACACATTTTGATTTCAGAGTCTCTTCCCT	2049
QY	1820	GATTATGTGCTCCAGAGATCTTCATTTCCACACCGCTGGAGATGCGGCTCACGGC	1879
Db	2050	GACTATACTCTCCAGAAATTTTCTCTTTGAAAGTACTACTGGAATTACATTGTATGGG	2109
QY	1880	ATGATCTACAAGCCCCACGCCCTTGACAGCCAGGAGAAAGACCCACCGTCTCTTTGTA	1939
Db	2110	ATGCTCTACAAGCCTCATGTATCTACAGCCGGAAGAAATATCTACTGTGTCTATCA	2169
QY	1940	TATGAGAGCCCCCAGAGTGCAGCTGTGTAATTACTCTTAAAGCATCAAGTACTTGGCG	1999
Db	2170	TATGTGTGCTCTCAGAGTGAGGTGGTAATATCCGTTTAAAGGAGTCAAGTATTTCCGC	2229
QY	2000	CTCAACACACTGGGCTCTCCGGGCTACGCCGATGTGTATTTGACAGGAGGGGCTCCGT	2059
Db	2230	TTGTAATTAACCTAGCTCTCTTAGGTATATGTGTGTGTATGTATGACAAACAGGGATCTGT	2289
QY	2060	CAGCGAGGCTTCGGTTTCTGAAAGGGGCCCTGAAAACCAATGGCGAGGTGAGATCGAG	2119
Db	2290	CACCGAGGGCTTAAATTTGAAAGCGCCTTTAAATATAAATGGTCAAAATAGAAATTTGAC	2349
QY	2120	GACCAAGTGGAGGGCCCGCAGTTCGAGGCCAGAGATAGGCTGTATGAGCTAGACCGGA	2179
Db	2350	GATTAAGTGGAGGACTCCATATCTTAGCTTCTCGATATGATTTATTAAGTACTTGTAGTGT	2409
QY	2180	GTTCGCATTCATGGCTGTGTCCTACAGGGGGCTTCCTTCGCTCATGTGGGCTATATCCACAG	2239
Db	2410	GTGGGCAATCCACGGCTGTCTCTATGAGAGGATACCTCTCCCTGATGGCATTAATGACAGG	2469
QY	2240	CCCCAGGTGTTCAAAGTGTGGCCATCGCGGGTGCCCGCGTCACCGTGTGATGGCTACAGAC	2299
Db	2470	TCAGATATCTCATAGGGGTGTCTATTTGCTGTGGGGCCCGCCAGTCACTCTGTGATCTTGTATGAT	2529
QY	2300	ACAGAGTACACTGAGCGCTCATATGAGACGTCCCTGAGAAACAAACACAGACGGCTATGAGGGCG	2359
Db	2530	ACAGAGATACGAGACTTATATGGGTCAACCTGACACAGATATGAACAGGCTATTAATCTTA	2589
QY	2360	GGTTCGTCGGCCCTGCACAGTGAAGAGTCGCCAATGAGCCCAACCGCTTCTTATCTCC	2419
Db	2590	GGATCTGTGGCATGACAAAGCAAAAGTTCCCTGTGAACCAAAATCGTTTACTGTCTTAA	2649
QY	2420	CACGGCTTCTCGAGCAAAACGTCGACTTTTTCACACAACTTCTCTGCTCCCACTG	2479
Db	2650	CATGTGTTCTCTGATGTGAGATATGCCATTTTGGACATATCCAGATATTTTCTCAGTATTTTTTA	2709
QY	2480	ATTCGACACAGAAACCTTTCACAGCTCCAGATCTACCCCAAGAGACACAGATATTCGC	2539
Db	2710	GTGAGGGCTGGAAAGCATATGTATTTACAGATCTATCTCTCAGAGAGACACAGCATTAAGA	2769
QY	2540	TGCCCCGAGTGGGGCGAGCACTATGAAGTACAGTGTGTGTCACATTTCTACAGAAATACCTC	2599
Db	2770	GTTCCTGAATTCGGGAGAACATTTATGAACATGCAATCTTTTGCACATACCTTCAAGAAACCTT	2829
QY	2600	TGAGC 2604	
Db	2830	GGATC 2834	

RESULT 10			
BC030688			
LOCUS	3130 bp	mRNA	linear
BC030688			
DEFINITION	Homo sapiens, Similar to dipeptidylpeptidase 8, clone MGC:26191		
			PRI 30-MAY-2002

ACCESSION	IMAGE:4822550, mRNA, complete cds.
VERSION	BC030688
KEYWORDS	BC030688.1 GI:21265132
SOURCE	MGC.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL	1 (bases 1 to 3130)
	Strausberg,R.
	Direct Submission
	Submitted (24-MAY-2002) National Institutes of Health, Mammalian
	Gene Collection (MGC), Cancer Genomics Office, National Cancer
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
	USA
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT	Contact: MGC help desk
	Email: cgabs-remail.nih.gov
	Tissue Procurement: Miklos Pakovits, M.D., Ph.D.
	cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shihrai
	Toshiyuki and Piero Carninci (RIKEN)
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: Institute for Systems Biology
	http://www.systemsbio.org
	contact: amadansystemsbiology.org
	Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha
	Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
FEATURES	Clone distribution: MGC clone distribution information can be found
source	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
	Series: IRAC Plate: 33 Row: d Column: 5
	This clone was selected for full length sequencing because it
	passed the following selection criteria: matched mRNA gi: 18450277.
	Location/unalifers
	1. .3130
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="MGC:26191 IMAGE:4822550"
	/tissue_type="Testis"
	/clone_id="NIH_MGC_97"
	/lab_host="DH10B"
	/note="Vector: pbluescript"
	372. .2867
	/codon_start=1
	/product="Similar to dipeptidyl peptidase 8"
	/protein_id="AAH30688.1"
	/db_xref="GI:21265133"
	/translation="MAAAMETDEGLVIFPTACCEENIESODRPKLEPFYVERYSQ
	LKRLADRTKRYHGMKAPHDEFAVRNDPDGPHSRITTLAMSGERNRNTLYSEI
	PTFTNRAAVLMLSWKPLDLFQATLDLGMVSRREELRLREKRLGVIGIADVDHQSG
	KTEFGSGGIYVNGDGGPOGFTQDPLRPNLVETSPNIRMDPKCADPDMWIAFHSN
	DIATSNIVTREERLTYVNHNLAMMEDASAGVAFVLDEEFDKSGYWKCAEFT
	PSGKILIRLIEENDESEVELIHTYSMLTFRADSRFRYKGTGANPKYVFKSEIMI
	DABRIITDIVIDKELIOPFELFBEVEYIARAAGTPEBKVAMSLILDSORTDIVLIS
	PELLEPVEDVMEKRLIESVPDSVTPLLIIEETDWINIHIFVFPDSHEELEF
	IFPAECKTFRHRLKITSILKESKYSKSSGLPASPFCPIKEELAITSGEWEVLGR
	HGSIIOVDEVARLVEYFEGTDSPLDEHLIYVSVNPEVPIRLDRGSHSCCISOHD
	FEIISKYNOKNPHCVSLYKLSPPEDDTGCKETEMAIILDSAGLPDLPYPELTFES
	TGTGTLGMLYKHPDLOPKKRYPLVLFVCGPOVOLYNNRFGVYKPRNLTAISLGY
	VVYDITGNSCHRGKLEFGARKYKMAVIAAGPVTIMLYFDGTGYTERVHGHPDNGCY
	LGVSAVMOAEKRPSEPNRLILHGFLENDVHFHFTSILSLFVLRAGKRYDQLYPOBRH
	SIRPESGEHYEHLILYLDENLGSRTAAALKVI"
BASE COUNT	955 a 626 c 717 g 832 t
ORIGIN	
Query Match	24.9%; Score 651.8; DB: 9; Length 3130;
Best Local Similarity	58.3%; Pred. No. 2.8e-101;
Matches 1180; Conservative 0; Mismatches 837; Indels 6; Gaps 2;	
07	80 TTCCAGGCGCAGAACACTCGCTGGAGCGGCTCCGAGACATCAACGAGCGCGGAG 139
474	TTTTATGTTGAGCGGATTCCTGAGAGTCAGCTTAAAAAGCTGCTGGCGGATACCAAGAAA 533

Kearney, L., Tribouley, C.M., Khan, F.A., Yao, M.G., Baughn, M.R.,
 Azimzai, Y., Elliott, V.S., Nguyen, D.B., Gandhi, A.R., Yang, J.,
 Hernandez, R., Pollock, J.L., Lu, D.A., Reddy, R., and Tang, Y.T.,
 Proteases
 Patent: WO 0183775-A 16 08-NOV-2001;

JOURNAL
 Incyte Genomics, Inc. (US)

FEATURES
 Location/Qualifiers

source
 1.2510
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 376067Cbl"
 BASE COUNT 777 a 494 c 527 g 712 t
 ORIGIN

Query Match 21.9%; Score 573.8; DB 6; Length 2510;
 Best local Similarity 57.4%; Pred. No. 5.4e-88;
 Matches 1138; Conservative 0; Mismatches 822; Indels 21; Gaps 5;

QY 127 CGGACCCGCACTACTCGGGCTCATTTGTCAACAAGCGGCCGACGCTTCCAGTTTGT 186
 DB 29 CGATTACAGAAATATCATGGCTACATGATGGCTAAGGCCACATGATTTTCATGTTTGT 88
 QY 187 GCAGAGAGAGATGATGCTGGGGCCACGCTCCGCTCTACTACTGGGATGSCATA 246
 DB 89 GAAGAGGAATGATCCAGATGGACCTCATTCAGACAGATCTATTCCTGCTGCTGG 148
 QY 247 TGGCAGCCGAGAGAACCTCCCTCTCTACTCTGAGATTTCCCAAGAGTCCGGAAGAGCG 306
 DB 149 TGAAGACAGAGAAATACACTGTTTATTTCTGAAATTTCCCAAAACATCATATGAGCAGC 208
 QY 307 TCTGCTGCTCTGCTGGAAGCAGATGCTGATTCATTTCCAGGCCACGCCACCATG 366
 DB 209 AGCTTAATGCTCTCTGGAAGGCTCTTTTGATTTTTCAGGCAACACTGAGACTATGG 268
 QY 367 GGTCTACTCGGAGAGAGAGCTGCTGAGGAGCGGAAAGCGCTGGGCTCTCGGCA 426
 DB 269 AATGATTTCTCGAAGAAAGAACTATTAAAGAAAGAAAGCGATTGGAACTCGGAAT 328
 QY 427 CACTCTCTAGATTCACAGCAGAGAGTGGCCCTCTCTCTTCCAGGCGCAACAGCT 486
 DB 329 TCGTTCTTGAATATTCACCAAGAGTGGAACTTTCTGTTTCAGCGCGGTATGGAA 388
 QY 487 CTTCACCTGCTCGGAGCGGCGCAAGAGCGCTTCATGCTGCTCCCTATGAAACCGCTGA 546
 DB 389 TTATCTACGTAAGAGATGAGGCGCCACAGAGATTTACGCAACACTTTAAAGGCCCATCT 448
 QY 547 AATCAGAGCCAGTGTCTCAGGCGCGGAGTGGACCCCAAAATGTCCTGCGGACCTGC 606
 DB 449 AGTGAATAGTATGTCACCAACATACGAGATGATCCAAATTAATGACCTGATCCAGA 508
 QY 607 CTCTCTCTCTCATCATATACAGCAGCAGCTGAGTGGCCCAACATGAGACAGAGCGAG 666
 DB 509 CTGAGATGCTTTATATCAGAGAGATTTTGGATATCTTAACATGTTAAACCGAGAA 568
 QY 667 GCGGCGGCTGACCTTCTGCCACCAAGGTTTATCAATGCTCGGATGATGCCCAAGTCTG 726
 DB 569 AAGGAGACTACCTATATGTACACATAGCTAGCCACATGAGAAAGATGCGAGATCACG 628
 QY 727 GGGTGGGCGCCTGCTCATACAGGAGAGATTCAGCCGCTTCACGGGATGAGTGG 786
 DB 629 TGGAGTGGCTACTTGTGTTCTCCAAAGAAATTTGATTAATTTGCGGCTATGGTGG 688
 QY 787 CCCACAGCCTCTGGGAGAGTTCAGAGGCGCTCAAGACGCTGCGAATCCTGATGAGA 846
 DB 689 TCCAAAG--CTGAACAACATCCAGTGGTGGTAAATTTCTTGAATCTTATGAAGA 745
 QY 847 AGTCATGATCCGAGGTGAGGTCAATTCAGTCCCTCTCTCGGCTGAGAGAAAG--A 905
 DB 746 AATGATGATGATGAGGTGAGAAATTTATCATGTTACATCCCTATTTGGAACAAGCA 805
 QY 906 AAGAGAGCTGATGAGTCCCGAGAGAGGAGCAAGAAATCCCAAGATTTGCTTAAAC 965
 DB 806 GGGCAGATTCATTCGTTATCTTAAGAGGTACAGCAAAATCCTAAAGTCACTTTAAGA 865

QY 966 TGGCTGATTTCCAGACTGACAGCCAGGCGAAGATGCTCTGACCCAGAGAAAGAGCTGG 1025
 DB 866 TGTACAGAAATATGATTTGATGCTGGAAGAGATCATATGATGATATGATATGAGAACTA 925
 QY 1026 TGCAGCCCTTCAGCTGCTGTTTCCCAAGGTGAGTACATGCCAGGCGCGGTGACCC 1085
 DB 926 TTCAACCTTTTGAATCTTATTTGAAGAGATTTGAATATATTTCCAGAGCTGATGACTC 985
 QY 1086 GGGATGGCAATATGCTGGGCGATGCTCTGACCGCGGCCGAGCTGGCTCAGCTCG 1145
 DB 986 CTGAGGGAATATGCTGCTGCTGCTCTACTAGATGCTCCGAGCTGCTCAATATG 1045
 QY 1146 TCTCTCTCCCGCGCGCTCTGTTTCATCCAGACAGAGATGAGAGAGCGCTAGCCT 1205
 DB 1046 TGTGATATCAGCTGATATTTATTTATCCAGTACAGATGATGTTATGGAAGCAGAGAC 1105
 QY 1206 CTGCGAGCTGTCCCAAGAGATGTCAGCCGATGCTGATGCTGATGAGAGATCAGAC 1265
 DB 1106 TCAATGAGTACAGCTGATGATTTCTGATGAGCGCACTAAATATCTATGAAGAAACAGACA 1165
 QY 1266 TCTGATGCAATGTCATGATCTCTATCCCTTCCCAATCAGAGGAGAGAGAGAGC 1325
 DB 1166 TCTGATGCAATGTCATGATCTCTATGATTTTCCCAAGCTCAGAGAG--GAAA 1222
 QY 1326 TCTGCTTCTCCGCGCAATGATGATGACAGCCGCTTCTGCTTGTATCAAAAGTACCG 1385
 DB 1223 TTGAGTTTATTTTGGCTGATGATGCAAAACAGTTTCCGTATTTATCAAAATTTACAT 1282
 QY 1386 CCGTTTAAATCCAGGCTTACAGTGTGAGTGGAGCCCTTCAGCCCCGGGGAAGATGAT 1445
 DB 1283 CTATTTTAA-----GGAAGCAAAATATAAGATTCAGTGGTGGGCTGCT 1332
 QY 1446 TTATGTCCTCATTAAGAGAGATGCTCTGACAGCGGTGATGAGAGTTTGGCA 1505
 DB 1333 CCACGTCACCTGATGATGATCATTATGAGATGATGATGATGATGATGATGATGAT 1392
 QY 1506 GGCAGGCTCA-----AGATCTGGGTCAATGAGAGACCAAGCTGGTACTTCCAGGCG 1561
 DB 1393 GTGACACATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1452
 QY 1562 ACCAGAGACGCGCTGAGACACACCTCTGATGATGATGATGATGATGATGATGATGAT 1621
 DB 1453 ACCAAAGACTCCCTTTAGAGCATCTGATGATGATGATGATGATGATGATGATGAT 1512
 QY 1622 ATGCTAGGCTCCAGCGCGCTCTCCATGATGATGATGATGATGATGATGATGATGAT 1681
 DB 1513 GTGACAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1572
 QY 1682 ATGCTAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1741
 DB 1573 TTCTTTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1632
 QY 1742 AGCGCCCGGAGAGACCCCTGCAACAGAGCCCGCTTCTGCGCTGATGATGATGAT 1801
 DB 1633 TCAAGTCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1692
 QY 1802 GAGAGAGCTGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1861
 DB 1693 TCAAGAGTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1752
 QY 1862 GATGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1921
 DB 1753 GATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1812
 QY 1922 CCACCGCTCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1981
 DB 1813 CTTACTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1872
 QY 1982 GGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2041
 DB 1873 GAGTCAATATTTCCGCTTGAATGATGATGATGATGATGATGATGATGATGATGAT 1932

QY 2042 GAGGCGAGGGCTCTGTCAGCGAGGGCTTCGGTTCGAGGGGCGCTGAAACCAATG 2101
111 111111 111111 111111 111111 111111 111111 111111 111111 111111
Db 1933 GACACAGGGGATCCCTCACCAGGGCTTAATTTAGAGCGCGCTTAATATATATAT 1992
QY 2102 G 2102
1
Db 1993 G 1993

RESULT 12
AX405770 2842 bp DNA linear PAT 14-JUN-2002
LOCUS Sequence 185 from Patent WO0222660.
DEFINITION AX405770
ACCESSION AX405770
VERSION AX405770.1 GI:21438979
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Tang,Y.T., Liu,C., Zhou,P., Asundi,Y., Zhang,J., Zhao,Q.A., Ren,F.,
Xue,A.J., Yang,Y., Wehrman,T., and Drmanac,R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 0222660-A 185 21-MAR-2002;
HYSEO, INC. (US)

FEATURES
Source location/Qualifiers
1..2842
/organism="Homo sapiens"
/db_xref="taxon:9606"
234..2582
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD34877.1"
/db_xref="GI:21438980"
/translation="MAAAMETBOLGVEIFETADCEENIESODRPKLEPFYEVERYSMSQ
LKLIADTRKHGYMMAKAPHDPMFVRNDPDSRIRIYAMSGENRENTLFYSI
PKTINAAVAVLMSKPLDFOATLIDGMSREBELREKRIGTGYIADYHOGSG
TEFPOAGSGYHYKDGSGPOGFOPLRPNVVERSCPIRNDRLCPADPWIAPFISH
DIWISNIVREERRLTYVHNLNMEEDASAVATVLOEEDRSIGYWCRAEET
PSGKILIRILENDESEVEIIVTSPMLTBRADSRFRYGTANRKYFKSEIMI
DAEGRILIDVDEKLIOPFELLEGVEIARAGWPEKYMSTILLDSOTRLOIVLS
PELIEVEDVMEORRLIESVPSVPLIYEETDIIIMIHIFVEPQSHHEIEIF
IFASECTGFRHLKYITSLIKESKYSRSGGLPAPSDKCPKEBELATISGMEVIGR
HGSNIQVDEYRRLVYRPGTDSPLHHLIYVSYNREYRLDRKSHSCCISOHD
FETSKYSNQNPRCVSLYKUSPEDDTCTKKEWATILDSAPLPDYTPETLFSES
TTFETLYGMLYKPHDLQPKRYPVLEIYGGPOVALAGAVTLMIEFYDTGTRYMGH
PDQNEGYLYGSVMAQAEKPESEPNRLILHLGFLDENVFAHTVLSILSFLVRAGKPYD
LQIYPOGRHSIRVPESEHVELHLHLHYLENLGSRIALKVI"

BASE COUNT 857 a 592 c 635 g 758 t
ORIGIN

Query Match 21.7%; Score 568.2; DB 6; Length 2842;
Best Local Similarity 57.2%; Pred. No. 4.8e-87;
Matches 1073; Conservative 0; Mismatches 798; Indels 6; Gaps 2;

Db 576 TCTTGAAGAGCTCTTTGGATCTTTTCAGGCAACACTGGACATATGGAATATATTCGA 635
QY 380 GAGAGAGAGCTGCTGAGGAGCGGAAACGCTGGGGGTCTTGCGATCACCCTCTACGAC 439
Db 636 GAAAGAACATATTAAGAGAAAGAAACGCAATTCGGAACAGTCGGAATTCCTTACGAT 695
QY 440 TTCACAGCGAGATGGCGCTTCCTCTCCAGGCGACGAACACGCTTCCTACATGTCGC 499
Db 696 TATCACCAGAGAGATGGAACATTTCTGTTTCAGCGCGGATGGAATTTATCAGCTAAA 755
QY 500 GAGCGCGCAAGAAAGCGCTTCATGCTGTCCCTTGAACCGCTGGAATCAAGACCAG 559
Db 756 GATGAGAGGGCCACAAGGATTTAGCAACACCTTTAAGGCCATTCATAGTGAACATAGT 815
QY 560 TGTCTAGGCGCCCGGATGAGACCCCAAAATCTGCCCTGCGACCTGCCCTTCCTTC 619
Db 816 TGTCCAAACATACGATGATGATCCAAAATTAATGCTCTGTATCCAGACTGGATTGCTTT 875
QY 620 ATCAATAAGCGACCTGTGGGTGGCCACATCGAGACAGGCGAGGCGGCGCTGACC 679
Db 876 ATACATGCAACGATATTTGATATCTTACATCTGTAACGAGAGAAAGAGACTACT 935
QY 680 TTCTGCCACCAAGGTTTATCCATGCTCCTGATGACCCCAAGTCTGGGGGTGGCCACC 739
Db 936 TATGTCACATGAGCTAGACCAACATGGAAGAAAGATGCAATCAGCTGGAGTGTAC 995
QY 740 TTCTCATATACGAAGAGTTCAGCCCTTCACTGGGTACTGGTGTGCCACAGCTCC 799
Db 996 TTCTTCTCCAAAGAAATTTGATATGATATCTGGCTATTTGGTGTGCCAAAAG---CT 1052
QY 800 TGGGAAGTTCCAGAGGGCTTCAAGCGCTGCAACGCTGATTCGTATGAGAAAGTATGAGTCC 859
Db 1053 GAAACACTCCCACTGGGTGATTAATTTCTTGAATTTCTATATGAGAAATGATGAATCT 1112
QY 860 GAGGTGAGAGCTATTCAGTCCCTCTCTGCTGCTGCTAGCAAGAAAGAGAGAGCTGAT 919
Db 1113 GAGGTGGAATATTTATCATGTTACATCCCTATGTTGGAAACAAAGAGGCGAGATTCAATC 1172
QY 920 CGGTACCCCGAGAGAGCAGCAAGAACATCCAGATTTGCTTGAATCTGCTGAGTCCAG 979
Db 1173 CGTATCTTAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1232
QY 980 ACTACAGCCAGGCGCAAGATGCTCGACCCAGAGAGAGAGAGTGTCCACCTTCAGC 1039
Db 1233 ATGATGCTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1292
QY 1040 TCGCTGTTCCGGAAGGTGAGTACATGCGCAGGCGGAGTGGACCCGAGATGCAAAATAC 1099
Db 1293 ATTCTATTTGAAGAGATGATGATATATTCAGAGCTGATGATGATGATGATGATGATGAT 1352
QY 1100 GCTTGGGCGATGTTCTGAGCGCGGCCCGCAGAGTGGCTCCAGCTGCTCTCCGCCCG 1159
Db 1353 GCTTGGTGCATCTTACTGATGATGCTGCTCCAGACTGCGCTACAGATAGTGTATCTACCT 1412
QY 1160 GCCCTGTTCACTCCGAGACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1219
Db 1413 GAATTAATTTATCCAGTAGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1472
QY 1220 CCCAGGATGTCACCGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1279
Db 1473 CCTGATTTGAGAGCGCCAT 1532
QY 1280 CATGACATCTTCTATCTCTCCCTCCCAATCAGAGGAGAGAGAGAGAGAGAGAGAGAG 1339
Db 1533 CATGACATCTTCTATCTTCTTCCCAAGAGTACAGAGAG---GAATTAATGATTAATTTT 1589
QY 1340 GCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1399
Db 1590 GCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1459
QY 1400 CAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1459

Db 1650 AGCAAAATATAAGCATCAGTGGTGGCTGCTGCTCCAAAGTGAATTCAGATGTCCTATC 1709
QY 1460 AAGGAAGAGATGCTCTGACACAGCGTGAATGGAGAGTTTGGCGAGCAGCGCTCCAG 1519
Db 1710 AAGAGAGATGATACATTTACAGAGTGGATGGAAAGTTCTTGGCCGCGCATGGATCTAAT 1769
QY 1520 ATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGGCACAAGACAGACCGCTG 1579
Db 1770 ATCCAAGTGTAGAGCACAAGAGCTGTGTATTTTGAAGCACAAGAACTCCCTTTTA 1829
QY 1580 GAGCACCACCTTACGCGGTGACGTATGAGCGCGCCGCGCAGATCGTACGCTCACCAG 1639
Db 1830 GAGCATCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1889
QY 1640 CCGGCTCTTCCATAGCTCTCCATGAGCCAGAACTTCAGATGTTCTGACGACCTAC 1699
Db 1890 CGTGGTACATCATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1949
QY 1700 AGCAGGTGAGCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1759
Db 1950 AGTAAACAGAGATCAGACACTGTGTGCTGCTTACAGCTATCAAGTCTCTGAAGATGAC 2009
QY 1760 CCGCTGACAGAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1819
Db 2010 CCACTTGTCAAAACAAAGAAATTTTGGCCACCACTTTTGGATTCAGCAGGCTCTTCT 2069
QY 1820 GATTATGTTCTCTCAGATCTTCCATTTCCACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1879
Db 2070 GACTATATCTCTCCAGAAATTTTCTTTTGAAGATCTCTGATTTACATTTGTATGGG 2129
QY 1880 ATGATCTACAGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1939
Db 2130 ATGCTACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2189
QY 1940 TATGAGAGCCCGCAGGT 1956
Db 2190 TATGTGTGCTCTCAGGT 2206

RESULT 13
AK000290
LOCUS AK000290 2778 bp mRNA linear PRI 22-FEB-2000
DEFINITION Homo sapiens CDNA FLJ20283 fls, clone HEP04088.
VERSION AK000290.1 GI:7020272
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:Hepeg2 CDNA to mRNA, clone_lfb:HEP
clone:HEP04088.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T., and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2778)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T., and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
source Location/Qualifiers
1..2778
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP04088"
/cell_line="Hepeg2"
/cell_type="hepatoma"
/clone_lfb="HEP"
/note="cloning vector pME18SFL3"
163..1572
/note="unnamed protein product"
/codon_start=1
/protein_id="BA91059.1"
/db_xref="GI:7020273"
/translation="MAAMETDGLVEIFETADCEENIESODRPLEPYVERYSMSO
LKLADPTKHYGWMKAPHDMPYKRPNDPPGSPDRYLLYLAAMSGENENTLFYSEI
PTKINAAVLMISMKPLDLPQATLDYGNATSHREELLRKRRTGYGTASTYDHOSSG
TFLFGSGITVHKDGGPOGFTQOPLRPNLVETSCPNIIMDPKLCFADPDWIAFHSN
DIWISNIVREERRLTYVHNEELAMBEDASGAVFVLOEEDRYSQWMCPRKATTT
PSGKILRLIYEENDESEVEIHTVSPMLETRADSPRYKGTGNAPKVTFMSEIMI
DAEGRIDVIDKELIQPELIFEGVEYIARAGWTPGKXYAMSLIDRSOTRQIYLIS
PELIFVEDDVMERORLIESVDPVPLIYEETDWMINDIHVFPQSHHELEIF
IFASEKTEFRHLIKITST"
BASE COUNT 856 a 562 c 615 g 745 t
ORIGIN
Query Match 21.3%; Score 557.8; DB 9; Length 2778;
Best Local Similarity 57.2%; Pred. No. 2.8e-85;
Matches 1073; Conservative 0; Mismatches 797; Indels 7; Gaps 3;
QY 80 TTCCAGGTGAGAGACGATCGTGGAGCGGCTCCGAGCATATCCACGAGCCGCAAG 139
Db 265 TTTTATGTGAGGGGTATTCCTCGAGTACACTTAAAGCTGCTGGCCATACCAAGAAA 324
QY 140 TACTCGGCTCATTTGTCAACAAGCGCCCGCAGACACTTCCAGTTTGTGCAAGAGCGAT 139
Db 325 TATCATGCGTACATGATGAGGCGTCAAGCGCACATGATGATGATGATGATGATGATGATGAT 384
QY 200 GAGTCTGGGCGCCACCTCCACCGCTCTACTACTGAGGATGATGATGATGATGATGATGATGATGAT 289
Db 385 CCAAGTGGACCTTATTCAGACAGAAATTTACTTCCATGCTGTGTGAGAACAGAGAA 444
QY 260 AACTCCCTCTTACTCTGAGATTTCCCAAGAGGCTCGGAAGAGAGCTGTGCTCTG 319
Db 445 AATACACTGTTTATTTCTGAATTTCCCAAAATATCAATGAGACAGACAGCTTAATGCTC 504
QY 320 TCTGTGAGAGAGATGCTGATTCATTTCCAGGCCACAGCCCGCCACCATGGGGTCTACTCGG 379
Db 505 TCTTGAAGCTCTTTTGGATCTTTTCAAGCAACTGGACATGATGATGATGATGATGATGATGATGAT 564
QY 380 GAGGAGAGCTGCTGAGAGGAGCGGAAACGCTGGGGGCTTTCGAGCATACCTCTACGAC 439
Db 565 GAGGAGAACTATTAGAGAGAAAGAAACGATTTGGAACATCGGAATGCTTCTTACGAT 624
QY 440 TTCCACAGCAGAGTGGCTCTTCTCTTCCAGGCCACAGCAACGCTCTTCCACTGTGCG 439
Db 625 TATCACCAAGAGATGAGAACATTTCTGTTCAAGCGGCTAGTGAATTTATCAAGTAAAA 684
QY 500 GAGCGGCGCAAGAACGCTTCATGATGTGCTCCCATGAAACCGCTGGAATCAAGACCCAG 559
Db 685 GATGAGAGGCGCACAGAGATTTACGCAACAACCTTTAAGGCGCAATCTAGTGAACATAGT 744
QY 560 TGCTCAGGGCGCCGAGTGAAGGAGCCCAAAATTTGCTGCGCAGCCAGCTCTCTCTCTTC 619
Db 745 TGTCCCAACATACGATGATGATTCAAAATTTATGCGCTGCTGATTCAGACTGATGATGATGATGAT 804
QY 620 ATCAATAACAGCACTGTGGTGGCCCAACATGAGACAGCAGAGAGCGCGCTGAC 679
Db 805 ATACATAGCAAGATATTGGATATGATATACATGCTGAACAGAGAAGAGAGACATCACT 864
QY 680 TTTCGCCACCAAGTTTATGCAATGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 739
Db 865 TATGTGCACAAATGAGCTAGCAACATGGAAGATGCCAGATCAGCTGATGCTGCTTACC 924

QY	1820	GATTATGTCCTCCAGAGATCTTCATTTCCACACAGCGCTCGAGTGGCGCTACAGCG	1879
Db	1998	GACATACCTCTCCAGAAATTTTCTCTTTTGGAAAGTACTACTGATTTACATTTGATGGG	2057
QY	1880	ATGATCTACAAAGCCCGACCGCTTGACAGCCAGGAGAAAGACCCCGCTCCTCTTTGTA	1939
Db	2058	ATGCTCTACAGAGCCCTCAGATCTACAGCGCTGGAAAGAAATATCCATACGTCGTTCATA	2117
QY	1940	TATGAGAGCCCCAGAGT	1956
Db	2118	TATGTGGTCTCCTCAGAGT	2134
RESULT 14			
AK027826			
LOCUS			
DEFINITION	AK027826	Homo sapiens cDNA FLJ14920 fls, clone PLACE1007416, weakly similar to D1EPED1DVL PEPTIDASE IV (BC 3.4.14.5).	
ACCESSION	AK027826		
VERSION	AK027826.1	GI:14042786	
KEYWORDS		oligo capping; fls (full insert sequence).	
SOURCE		Homo sapiens placenta cDNA to mRNA, clone_1lb:PLACE1 clone:PLACE1007416.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS		1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shitatori,A., Sudo,H., Magatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuo,Y., Ninomiya,K. and Watanagi,T.	
TITLE		NEDO human cDNA sequencing project	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 2161)	
AUTHORS		Isogai,T. and Otsuki,T.	
TITLE		Direct Submission	
COMMENT		Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory: 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.	
FEATURES			
SOURCE		Location/Qualifiers	
		1..2161	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="PLACE1007416"	
		/tissue_type="placenta"	
		/clone_1lb="PLACE1"	
		/note="cloning vector: PME18SF3"	
		16..1914	
		/note="unnamed protein product"	
		/codon_start=1	
		/protein_id="BAB5395.1"	
		/db_xref="GI:14042790"	
		/translation="MDPKLCPADPDWIAFIHSNDIWSINITYRRERLTYVHNLAMNM EEARAGVATFVYLOEEFDRYSGYWMCPKAEPTSGSKILIRLIEENDESVELIHYLT SPLETPRARSFPRKGTGTANPKYTFKMSSEIMDADEGRIIDVIDKELLOPELIFEGV EYIARAKWTPDEKXAWSILDRSOTRLQIVLISELTPVEDDWEKORLIESVPDSV TPPLIIEETDIDWINIHIDHVPQSHSEELFTFAECKTGFRHLKRTISLSEKXY KRSGGIPAPDEKCPILKEEIAIHFAGSEWYVGRGNSINIQVDEARLYVPGTKSPLE HHLYVYVNPGEVTRLIDRSYSHSCISIOHCDGFIKSYSKNPKCVSKYLSLSPED DPCKCTEWPATLIDASGLPDYDYPPELIFESSTGFTGLGMKLPKHDLOGKKYPTV LFTYGGQVOLDVNNRKFQVYFRNLITASLGYVYVVDNKGSCRGILKFSBARTKAY AIVAGPITMIFDTQDITGTYTERKHPDNOEGGYIGVSAWAEKRPSPNNRLILHGLF DEWHPAHTSITLISLFEIRAKRPVDLQIYPERHSIRPEGEYEHHLIHYLDENLSS	
CDS			

IFASECKTGRHLKYITSLIKESKYNRSSGGLPAPSDFKPIKEEIAITSGEWEVLGR
HGSNIQVDEARLRYVEEGTDEPLEHLLVYSVYNVEVRLDRGSHSCJISOHD
FETSKXSNOKNPHVCEISLYKLSPEDEPTCKREPMATILDSAPLDYRPPLEFSEPS
TTFEFLXGMLXKPHDLOPKKYPVLFYGGPVAIAGATVITLMIPIDYTERYKMH
PDNEDGYLIGVAMQAEKPESEPNRLHLHGLDEVDENFAHRSILSLFLVRAGKPYD
LQIYPERHSIRPESGEHEHLHLHYLDONILSRIALAKVI*

BASE COUNT 796 a 564 c 592 g 716 t
ORIGIN

Query Match 16.6%; Score 434.6; DB 6; Length 2668;
Best Local Similarity 58.1%; Pred. No. 2,7e-64;
Matches 805; Conservative 0; Mismatches 574; Indels 6; Gaps 2;

QY 572 CGAGTGGACCCCAAAATCTGCGCCGACCCCTTCTCTCTCATCAATAACAGC 631
DB 654 CGGATGGATCCAAATTTATGCGCTGATGCAGACTGGATTGCTTTATACATAGCAAC 713
QY 632 GACCTGTGGGTGGCCCAATCGAGACGCGGCGGCTGACCTTCTGCCACCAA 691
DB 714 GATATTTGGATATCTAAGCATGTAACCAAGAGAAGAAAGAGACTATGTGCACAAAT 773
QY 692 GGTTTATCCAAATGTCCTGGATGACCCCAAGTCTCGGGTGTGGCCACCTTCTCATACAG 751
DB 774 GAGCTACCCCAACATGGAAGAGATGCCAGATCAGCTGGAGTGGCTACCTTTGTTCTCCAA 833
QY 752 GAAGAGTTGCGCCGCTCAGTGGTACTGGTGGTGGCCCAAGGCTCTCGGGAAGTTCA 811
DB 834 GAAGAAATTTGATGATATTTCTGGCTATTTGGTGGTCCAAAG--CTGAACAACATCC 890
QY 812 GAGGCGCTCAAGACGCTGCAATCTGTATGAGGAAGTCATGATGTCGAGGTGAGGTC 871
DB 891 AGTGGTGGTAAATTTCTAGATTTCTATATGAAGAAATGATGATCTGAGAGTGAATTT 950
QY 872 ATTCAAGTCCCCCTCTCTGCGCTAGAAAGAAAGAAAGAGAGCTGTATGCTGACCCAGG 931
DB 951 ATTCAATGTTACATCCCTATGTTGAAACAAGAGAGGAGATTCATTCCTTATCTCAA 1010
QY 932 ACAGGACGCAAGATCCCAAGATGCGCTGAACCTGGTGGTTCGACAGCAGACGAG 991
DB 1011 ACAGGTACAGCAAAATCTTAAGTCACTTTTAAGATGTCGAATTAATGATGATGCTGAA 1070
QY 992 GGCAGATTCGCTCTGACCCAGAGAGAGAGAGCTGTGACGCTTCAAGCTGCTGTTCCCG 1051
DB 1071 GGAAGATTCATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1130
QY 1052 AAGGTGAGTACATGCGCAGGCGCGGCTGACCCGGGATGCAATATACGCTGGGCCATG 1111
DB 1131 GGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1190
QY 1112 TTCTTGACCGCGCCCAAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1171
DB 1191 CTACTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1250
QY 1172 CCGAGACAGAGATGAG 1231
DB 1251 CCACTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1310
QY 1232 CAGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1291
DB 1311 ACGCCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1370
QY 1292 TATCCCTTCCCGCAATGAG 1351
DB 1371 CATGTTTTTCCCAAGAGTCAAGAGAG--GAAATTAATTAATTAATTAATTAATTAATTA 1427
QY 1352 AAGACCGGCTTCTGCAATTTGTACAAAGTCAACCCGCTTTTAAATCCAGGGCTACGAT 1411
DB 1428 AAAACAGGTTTCCGTCATTTATACAAATTTATACATCTATTTTAAAGAAAGCAATATATA 1487
QY 1412 TGAAGTACGCTTCAAGCCCGGGAAGAGATTAATTAATTAATTAATTAATTAATTAATTA 1471
DB 1488 CGATTCAGTGGTGGGCTGCTCAAGATGATTTCAAGTGTCTCTATCAAGAGAGAGATA 1547

QY 1472 GCTCTGACACGCGTGAATGGAGGATTTTGGCGAGGACGAGGCTCCAGATCTGGGTCAAT 1531
DB 1548 GCAATTAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1607
QY 1532 GAGAGACCAAGCTGTGTACTTCCAGGGCACCAAGAGACGCGCTGGAGACACACCTC 1591
DB 1608 GAACTCAAGAGGCTGTGTATTAATTTGAAGGCACCAAGAGCTCCCTTTAAGACATACCTG 1667
QY 1592 TACGTGTACACTATGAGGCGCGCGGAGATCTGAGCCCTACACAGCCCGGCTTCTCC 1651
DB 1668 TACGTGTACACTATGAGGCGCGCGGAGATCTGAGCCCTACACAGCCCGGCTTCTCC 1727
QY 1652 CATGAGTGTCTCATGAGGACCAAGCTGACATGTTGTCAGCCCACTACAGCAGCGTGAGC 1711
DB 1728 CATCTGTCTCATGAGGACCAAGCTGACATGTTGTCAGCCCACTACAGCAGCGTGAGC 1787
QY 1712 AGCGCGCCTGCTGACAGCTTACAGCTGAGCGGCCCGGACGAGACGACCCCTGCACAA 1771
DB 1788 AATTCACAGCTGTGTGCTCTTACAGCTATCAAGTCTGAAGATGACCCAACTTGCAA 1847
QY 1772 CAGCCCGCTTCTGTGGCTAGCATGATGAGGACGAGCAGCTCCCGGATTAATGTTCT 1831
DB 1848 ACAAGGAATTTTGGCCACCATTTTGGATTACAGAGGTCTCTTCTGACTATATCTCT 1907
QY 1832 CCAGGATCTTCCATTTCCACACGCGGCTGGATGCGGCTCTAAGGATGATCTACAA 1891
DB 1908 CCAGGAATTTCTCTTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1967
QY 1892 CCCCAGCCTTGCAGCGGAG 1951
DB 1968 CCTCATGATCTACAGCTGGAAGAAATATCTCTACTGTGCTGCTCATATATGATGATGAT 2027
QY 1952 CAGGT 1956
DB 2028 CAGGT 2032

Search completed: December 12, 2002, 09:56:58
Job time : 4513 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 07:16:06 : Search time 360 Seconds
(without alignments)
16370.777 Million cell updates/sec

Title: US-09-976-674-4

Perfect score: 2617
Sequence: 1 caagctaccatgcgcacaca.....ctcgagcgccgcgcatccg 2617

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

N.Geneseq-101002:*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2617	100.0	2617	24	ABK83323
2	2592.4	99.1	4219	24	ABK83335
3	2592.4	99.1	4302	24	ABK83333
4	2587.6	98.9	3024	24	ABK83339
5	2504.4	95.7	4180	24	ABK83338
6	2504.4	95.7	4263	24	ABK83338
7	2488.6	95.1	2495	24	ABK83337
8	2296.4	87.7	4076	24	ABK83337
9	2296.4	87.7	4159	24	ABK83336

10	2288.2	87.4	2801	22	AA157896	Human polynucleoti
11	2208.4	84.4	4037	24	ABK83341	CDNA encoding huma
12	2208.4	84.4	4120	24	ABK83340	CDNA encoding huma
13	2091.4	79.9	3262	22	AA157880	Human polynucleoti
14	1936.4	74.0	3287	24	AAK83855	Alternative versio
15	1869.6	71.4	2461	21	AACT5835	Human ORF1390
16	1867.4	71.4	2751	24	AAK83311	Murine dipeptidyl
17	1850.4	70.7	2982	22	AA159666	Human polynucleoti
18	1585.6	60.6	3047	24	ABK69113	DNA encoding human
19	1584.4	60.5	2952	24	ABK69090	DNA encoding human
20	1429.8	54.6	2411	24	ABK83334	CDNA encoding huma
21	1426.6	54.5	2079	21	AAA37672	Human peptidase, H
22	1386.2	53.0	2034	22	AA159682	Human polynucleoti
23	1316	50.3	2027	24	ABK69114	DNA encoding human
24	1308	50.0	2027	21	AACT7137	Human ORF1390
25	862.6	33.0	2643	24	AAH99335	Coding sequence of
26	862.6	33.0	2671	24	ABK83322	CDNA encoding huma
27	862.6	33.0	3106	24	ABK12892	Human protease PR
28	862.6	33.0	3143	24	AAH99334	CDNA encoding 2195
29	861	32.9	3120	22	AAK85694	Nucleotide sequenc
30	861	32.9	3120	24	AAK83856	Human dipeptidyl P
31	850.6	32.5	4829	24	ABK83327	CDNA encoding huma
32	658.2	25.2	4685	24	ABK83332	CDNA encoding huma
33	614.6	23.5	4676	24	ABK83331	CDNA encoding huma
34	573.8	21.9	2510	24	AAK23843	Human protease PR
35	568.2	21.7	2842	24	AAH59774	Novel human coding
36	566.4	21.6	4523	24	ABK83325	CDNA encoding huma
37	563	21.5	1048	22	AAK41004	CDNA encoding nove
38	518	19.8	662	22	AAK92083	Human CDNA clone r
39	518	19.8	662	22	AAK93366	Human full-length
40	518	19.8	1748	22	AAK94819	Human CDNA sequenc
41	516.6	19.7	2161	22	AAH15009	Novel human coding
42	445.6	17.0	4309	24	ABK83328	CDNA encoding huma
43	430.4	16.6	2668	24	ABH59775	Novel human coding
44	394.6	14.9	3713	23	ABH10425	Drosophila melanog
45	390.4	14.9	3783	23	ABH06641	Drosophila melanog

ALIGNMENTS

RESULT 1	
ID	ABK83323 standard; CDNA; 2617 BP.
XX	
AC	ABK83323:
XX	
DT	12-AUG-2002 (first entry)
XX	
DE	CDNA encoding human DPPIV related serine protease DPP-2.
XX	
KW	Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
KW	DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW	dyskinesia; reproductive disorder; inflammatory disorder;
KW	metabolic disorder; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200231134-A2.
PD	18-APR-2002.
XX	
PF	12-OCT-2001; 2001WO-US31874.
XX	
PR	12-OCT-2000; 2000US-240117P.
XX	
PA	(FERR) FERRING BV.
XX	
PI	Ol S, Akinsanya KO, Riviere PJ, Juntlen J;
XX	

DR MPI: 2002-444178/47.
DR P-PSDB: AB651592.
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX -
XX Claim 1: Page 56-57; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPV)-related
CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.
XX
SQ Sequence 2617 BP; 558 A; 830 C; 736 G; 493 T; 0 other:
Query Match 100.0%; Score 2617; DB 24; Length 2617;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 CAAGCTTACCATTGCGACACCGGGACCCCAACGCGCGACCGAGCGACCGCCAC 60
DB 1 CAAGCTTACCATTGCGACACCGGGACCCCAACGCGCGACCGAGCGACCGCCAC 60
OY 61 AGATGACCGCGCGCGCGCTTCCAGGTGCAAGACACTGCTGGAGCGGCTCCGGACAT 120
DB 61 AGATGACCGCGCGCGCGCTTCCAGGTGCAAGACACTGCTGGAGCGGCTCCGGACAT 120
OY 121 CATCCAGCGCAGCGCAAGTACTCGGGCTCATATTGCAACAMAGCGCCCGACACTTCCA 180
DB 121 CATCCAGCGCAGCGCAAGTACTCGGGCTCATATTGCAACAMAGCGCCCGACACTTCCA 180
OY 181 GTTTGTGCAAGAACGAGTGTGGGCCCACTCCACCGGCTTACTACTGGGAAT 240
DB 181 GTTTGTGCAAGAACGAGTGTGGGCCCACTCCACCGGCTTACTACTGGGAAT 240
OY 241 GCCATTGCGACGCGGAGAACTCCTCTACTCTGAGATTCCCAAGAAAGTCCGGAA 300
DB 241 GCCATTGCGACGCGGAGAACTCCTCTACTCTGAGATTCCCAAGAAAGTCCGGAA 300
OY 301 AGAGGCTGTGCTGCTCTCTGTAAGACAGATGTGATCATTTTCAGGCCACGCCCA 360
DB 301 AGAGGCTGTGCTGCTCTCTGTAAGACAGATGTGATCATTTTCAGGCCACGCCCA 360
OY 361 CCATGGGCTTACTCTCGGAGAGAGAGCTGTGAGGAGCGGAAAGCCCTGGGGTCTT 420
DB 361 CCATGGGCTTACTCTCGGAGAGAGAGCTGTGAGGAGCGGAAAGCCCTGGGGTCTT 420
OY 421 CGGCATCACCCTCTAGATTTCCACAGCAGAGTGGCTTCTTTCAGGCCAGCAA 480
DB 421 CGGCATCACCCTCTAGATTTCCACAGCAGAGTGGCTTCTTCTTCCAGGCCAGCAA 480
OY 481 CAGGCTCTTCCACTGTGCGAGCGGCGGAAAGAGGCTTCATGTTGCCCTTGAAC 540
DB 481 CAGGCTCTTCCACTGTGCGAGCGGCGGAAAGAGGCTTCATGTTGCCCTTGAAC 540
OY 541 GCTGGAATCAAGACCACTGTCAAGGCCCGGATGAGACCCCAAAATCTGCCCTGCA 600
DB 541 GCTGGAATCAAGACCACTGTCAAGGCCCGGATGAGACCCCAAAATCTGCCCTGCA 600
OY 601 CCTGCTCTTCTCTTATCAATAACAGCACTGTGGGTGGCCAAATGAGACAG 660
DB 601 CCTGCTCTTCTCTTATCAATAACAGCACTGTGGGTGGCCAAATGAGACAG 660

DB 601 CCTGCTCTTCTCTTATCAATAACAGCACTGTGGGTGGCCAAATGAGACAG 660
OY 661 CGAGAGCGCGGCTGACCTTGTGCCACCAAGTTTATCCAAATGTCTGGATGACCCAA 720
DB 661 CGAGAGCGCGGCTGACCTTGTGCCACCAAGTTTATCCAAATGTCTGGATGACCCAA 720
OY 721 GTCTGGGGGTGGCCACTTGTCAATACAGAAAGTTTCAGCGCTTCACTGGTACTG 780
DB 721 GTCTGGGGGTGGCCACTTGTCAATACAGAAAGTTTCAGCGCTTCACTGGTACTG 780
OY 781 GTGTGCCCCACAGCCCTCTGGGAAGTTTCAGAGGCGCTCAACAGCTGGATTCCTGTA 840
DB 781 GTGTGCCCCACAGCCCTCTGGGAAGTTTCAGAGGCGCTCAACAGCTGGATTCCTGTA 840
OY 841 TGAAGATGATGATTCGAGAGGTGAGGTCATTCAGTCCCTCTCTGCGCTAAGA 900
DB 841 TGAAGATGATGATTCGAGAGGTGAGGTCATTCAGTCCCTCTCTGCGCTAAGA 900
OY 901 AAGGAAGCGAGCTGTATCGGTACCCAGAGACGAGCAGCAAGATCCCAAGATTGCC 960
DB 901 AAGGAAGCGAGCTGTATCGGTACCCAGAGACGAGCAGCAAGATCCCAAGATTGCC 960
OY 961 GAACCTGGGTGATTCGACAGCTGACAGCCAGGCGCAAGATGCTGACCCAGGAAAGA 1020
DB 961 GAACCTGGGTGATTCGACAGCTGACAGCCAGGCGCAAGATGCTGACCCAGGAAAGA 1020
OY 1021 GCTGTGTACAGCCCTTACAGCTGCTGTTCGCCGAAGGTGAGTACATCGCCAGGCGG 1080
DB 1021 GCTGTGTGTACAGCCCTTACAGCTGCTGTTCGCCGAAGGTGAGTACATCGCCAGGCGG 1080
OY 1081 GACCGGGATGCAATATGCGCTGGGCCATGTTCTTGAGACCGGCGCCAGCATGGCTCCA 1140
DB 1081 GACCGGGATGCAATATGCGCTGGGCCATGTTCTTGAGACCGGCGCCAGCATGGCTCCA 1140
OY 1141 GCTGTGCTCTCTCCCGCGGCGCTGTTCATCCCGACACAGAAATGAGAGAGCGGCT 1200
DB 1141 GCTGTGCTCTCTCCCGCGGCGCTGTTCATCCCGACACAGAAATGAGAGAGCGGCT 1200
OY 1201 AGCCTCTGCGCAGAGCTGTGCTCCAGAAATGTCAGCCGTATGTGGTATGAGAGAGTCA 1260
DB 1201 AGCCTCTGCGCAGAGCTGTGCTCCAGAAATGTCAGCCGTATGTGGTATGAGAGAGTCA 1260
OY 1261 CAAGCTGTGATCAATGTTTCATGATCTTCTTCCCTTCCCAATCAGAGGAGAGA 1320
DB 1261 CAAGCTGTGATCAATGTTTCATGATCTTCTTCCCTTCCCAATCAGAGGAGAGA 1320
OY 1321 CGAGCTGTGCTTCTCCGCGGCCAATGAATGCAAGACCGGCTGTGCAATTTGACAAAG 1380
DB 1321 CGAGCTGTGCTTCTCCGCGGCCAATGAATGCAAGACCGGCTGTGCAATTTGACAAAG 1380
OY 1381 CACCGCGCTTTTAAATCCAGAGGCTACGATTGAGTGAGCCCTTTCAGGCCCGGGAGA 1440
DB 1381 CACCGCGCTTTTAAATCCAGAGGCTACGATTGAGTGAGCCCTTTCAGGCCCGGGAGA 1440
OY 1441 TGAATTTAAGTCCCATTAAGAGAGATGTGCTTGACACGCGTGAATGGAGGTTTT 1500
DB 1441 TGAATTTAAGTCCCATTAAGAGAGATGTGCTTGACACGCGTGAATGGAGGTTTT 1500
OY 1501 GGGAGAGCAAGGCTCAATGCTGGGTCAATGAGAGACCAAGCTGCTACTTCCAGGG 1560
DB 1501 GGGAGAGCAAGGCTCAATGCTGGGTCAATGAGAGACCAAGCTGCTACTTCCAGGG 1560
OY 1561 CACCAAGGACAGCGCGCTGAGACACCACTTACAGGTGACGTATGAGGCGCGGCA 1620
DB 1561 CACCAAGGACAGCGCGCTGAGAGACCACTTACAGGTGACGTATGAGGCGCGGCA 1620
OY 1621 GATGTAAGCTTACACAGCGCGGCTTTCCTCATAGCTGCTCCATGAGCCAGAACTTCA 1680
DB 1621 GATGTAAGCTTACACAGCGCGGCTTTCCTCATAGCTGCTCCATGAGCCAGAACTTCA 1680
OY 1681 CATGTTGCTGACGACTATACAGAGCGGTGAGACGCGGCTGCTGCTACAGCTCAAGCT 1740
DB 1681 CATGTTGCTGACGACTATACAGAGCGGTGAGACGCGGCTGCTGCTACAGCTCAAGCT 1740

QY	1741	GAGGGGCCCCGAGAGAGACCCCTCGACAAAGAGACCCCGCTTCTGGGTACGATGATGGA	1800
Db	1741	GAGGGGCCCCGAGAGAGACCCCTCGACAAAGAGACCCCGCTTCTGGGTACGATGATGGA	1800
QY	1801	GGAGCGCCAGCTGGCCCCCGGATTTATTTCTCTCCAGAGATCTTCCATTCCACACGGCTC	1860
Db	1801	GGAGCGCCAGCTGGCCCCCGGATTTATTTCTCTCCAGAGATCTTCCATTCCACACGGCTC	1860
QY	1861	GGATGTGCGGCTCTACGGCATGTATTACAAAGCCCCACGCTTTGCAGCCAGGGAAGAAACA	1920
Db	1861	GGATGTGCGGCTCTACGGCATGTATTACAAAGCCCCACGCTTTGCAGCCAGGGAAGAAACA	1920
QY	1921	CCCCACGCTCTCTTTGTATATGAGAGCCCCCAGAGGACGCGTGGATTAATCTCTTCAA	1980
Db	1921	CCCCACGCTCTCTTTGTATATGAGAGCCCCCAGAGGACGCGTGGATTAATCTCTTCAA	1980
QY	1981	AGGCATCAAGTACTTGGGGCTCAACACACTGGGCTTCCCTGGGCGCTACGCGTGGTGTGAT	2040
Db	1981	AGGCATCAAGTACTTGGGGCTCAACACACTGGGCTTCCCTGGGCGCTACGCGTGGTGTGAT	2040
QY	2041	TGACGGCAGGGGCTCTGTACGAGGAGGCTTGCGTTTCGAAGGGGCCCTGAAAAACCAAT	2100
Db	2041	TGACGGCAGGGGCTCTGTGTACGAGGAGGCTTGCGTTTCGAAGGGGCCCTGAAAAACCAAT	2100
QY	2101	GGGGCAGGTGGAGATCGAGGACCAAGGTGGAGGGGCTTCGAGTTTCGTGGCCGAGAAATAGG	2160
Db	2101	GGGGCAGGTGGAGATCGAGGACCAAGGTGGAGGGGCTTCGAGTTTCGTGGCCGAGAAATAGG	2160
QY	2161	CTTCAATGACCTGAGCGGATTTGCCATTGCGATGGCTGTGCTACGGGGGCTTCTCTCGCT	2220
Db	2161	CTTCAATGACCTGAGCGGATTTGCCATTGCGATGGCTGTGCTACGGGGGCTTCTCTCGCT	2220
QY	2221	CATGGGCTTAATCCACAAGCCCCAGGTGTTCAAAGGTGGCCATCGGGGTGCCCGGCTAC	2280
Db	2221	CATGGGCTTAATCCACAAGCCCCAGGTGTTCAAAGGTGGCCATCGGGGTGCCCGGCTAC	2280
QY	2281	CGTGTGATGGCCCTACAGACAGAGGATACATGAGGCTCATGAGACGTCCTGAGAAACAA	2340
Db	2281	CGTGTGATGGCCCTACAGACAGAGGATACATGAGGCTCATGAGACGTCCTGAGAAACAA	2340
QY	2341	CCAGCAGCGGTATGAGGCGGGGCTTCGTGGCCCTGACAGGTGAGAAAGTGGCCATGAGCC	2400
Db	2341	CCAGCAGCGGTATGAGGCGGGGTTCCGTGGCCCTGACAGGTGAGAAAGTGGCCATGAGCC	2400
QY	2401	CAACCGGTGCTTATCCTCCAGGCGTCTCGAGCGAAACGTGCACTTTTTCACACACAA	2460
Db	2401	CAACCGGTGCTTATCCTCCAGGCGTCTCGAGCGAAACGTGCACTTTTTCACACACAA	2460
QY	2461	CTTCTCTGCTCCCAACTGATTCGACAGGAAACCTTACACAGCTCCAGATCTACCCCA	2520
Db	2461	CTTCTCTGCTCCCAACTGATTCGACAGGAAACCTTACACAGCTCCAGATCTACCCCA	2520
QY	2521	CGAGAGACACAGTATTCGCTGCCCCCAGTGCAGGACGACTATGAAATCAGTGTCTGCA	2580
Db	2521	CGAGAGAGACAGTATTCGCTGCCCCCAGTGCAGGACGACTATGAAATCAGTGTCTGCA	2580
QY	2581	CTTCTACAGGAATACCTCTGACGGCCCGCGATCCG 2617	
Db	2581	CTTCTACAGGAATACCTCTGACGGCCCGCGATCCG 2617	
RESULT 2			
ABK83335			
ID	ABK83335 standard; cDNA; 4219 BP.		
XX	ABK83335;		
XX	12-Aug-2002 (first entry)		
XX	cDNA encoding human DPRP-2 splice variant #3.		
XX	Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;		

Query Match	Best Local Similarity	Matches 2593	Conservative	Score 2592.4	Pred. No. 0	DB 24	Length 4219
11	ATGGCCACCACCGGGACCCCAACGCGCCGACGAGCGGACCGCCCAATGACCCG	99.1%	0	1	0	0	0
367	ATGGCCACCACCGGGACCCCAACGCGCCGACGAGCGGACCGCCCAATGACCCG	100.0%	0	1	0	0	0
71	GGCGCCGCTTCAGATGTCGAGCAAGCACTGTGGAGCGGCTCCGGAGCATTCACGGC	99.1%	0	1	0	0	0
427	GGCGCCGCTTCAGATGTCGAGCAAGCACTGTGGAGCGGCTCCGGAGCATTCACGGC	99.1%	0	1	0	0	0
131	AGCCGCAAGTACTCGGGCTCATTTGTCACAAAGAGCGGCCACGACTTCCAGTTTGTCAG	99.1%	0	1	0	0	0
487	AGCCGCAAGTACTCGGGCTCATTTGTCACAAAGAGCGGCCACGACTTCCAGTTTGTCAG	99.1%	0	1	0	0	0
191	AGCGAGGATGATCTGGGCGCCCACTCCACGCGCTCTACTACCTGGGAATGCCATATGGC	99.1%	0	1	0	0	0
547	AGCGAGGATGATCTGGGCGCCCACTCCACGCGCTCTACTACCTGGGAATGCCATATGGC	99.1%	0	1	0	0	0
251	AGCGAGGATGATCTGGGCGCCCACTCCACGCGCTCTACTACCTGGGAATGCCATATGGC	99.1%	0	1	0	0	0
607	AGCGAGGATGATCTGGGCGCCCACTCCACGCGCTCTACTACCTGGGAATGCCATATGGC	99.1%	0	1	0	0	0
311	CTGCTCTGTCCTCGGAGAGCAGATGCTGATCATTTTCAGAGCCACGCGCCCAATGGGCTC	99.1%	0	1	0	0	0

|||||
Db 667 CTCTCTCTGTCGGAAGCAGATGCTGGATCATTTCCAGGCCACAGCCCCACCAAGGGGTC 726
QY 371 TACTCTGGAGAGAGAGCTGCTGAGGGAGCGGAACGCCCTGGGGGCTCTCGGCATACC 430
|||||
Db 727 TACTCTGGAGAGAGAGCTGCTGAGGGAGCGGAACGCCCTGGGGGCTCTCGGCATACC 786
QY 431 TCCTAGACTCTTCACACGAGAGTGGCCCTCTCTCTTCAGAGCCAGCAACAGCCTCTTC 490
|||||
Db 787 TCCTAGACTCTTCACACGAGAGTGGCCCTCTCTCTTCAGAGCCAGCAACAGCCTCTTC 846
QY 491 CACTGTGCGACGGCGGCAAGAGCGCTTATGCTGTCCCTATGAACCGCTGGAATC 550
|||||
Db 847 CACTGTGCGACGGCGGCAAGAGCGCTTATGCTGTCCCTATGAACCGCTGGAATC 906
QY 551 AAGACCCAGTGTCAAGGGCCCGGATGGACCCCAAAATCTGCCCTGGCACCCCTGTC 610
Db 907 AAGACCCAGTGTCAAGGGCCCGGATGGACCCCAAAATCTGCCCTGGCACCCCTGTC 966
QY 611 TTCTCTTCATCAATAACAGCAGCTGTGGTGGCCAACTCGAGACGAGAGAGCG 670
Db 967 TTCTCTTCATCAATAACAGCAGCTGTGGTGGCCAACTCGAGACGAGAGAGCG 1026
QY 671 CGGCTGACCTTCTGCCACCAAGGTTTATCCAATGCTCTGATGACCCCAAGTCTGGGT 730
Db 1027 CGGCTGACCTTCTGCCACCAAGGTTTATCCAATGCTCTGATGACCCCAAGTCTGGGT 1086
QY 731 GTGGCCACCTTCGTATACAGGAAGTTGACGGCTTCATCGGTTCTGCTGGTGGCCCC 790
Db 1087 GTGGCCACCTTCGTATACAGGAAGTTGACGGCTTCATCGGTTCTGCTGGTGGCCCC 1146
QY 791 ACAGCCTCTCTGGGAAGGTTTCAAGAGGGCCTCAAGACGCTGGCAATCTGTATGAGGAATC 850
Db 1147 ACAGCCTCTCTGGGAAGGTTTCAAGAGGGCCTCAAGACGCTGGCAATCTGTATGAGGAATC 1206
QY 851 GAGAGTCCGAGTGGAGGTCATTCACGTCCCTCTCTCTCGCTAGAGAAAGGAAGACG 910
Db 1207 GAGAGTCCGAGTGGAGGTCATTCACGTCCCTCTCTCTCGCTAGAGAAAGGAAGACG 1266
QY 911 GACTCGATGGTACCCAGAGCAGGAGCAAGATCCCAAGATTTGCTTGAACCTGGCT 970
Db 1267 GACTCGATGGTACCCAGAGCAGGAGCAAGATCCCAAGATTTGCTTGAACCTGGCT 1326
QY 971 GAGTCCAGACTACACGACGAGGCGAAGATGCTCTGACCCAGAGGAAGAGCTGTGACG 1030
Db 1327 GAGTCCAGACTACACGACGAGGCGAAGATGCTCTGACCCAGAGGAAGAGCTGTGACG 1386
QY 1031 CCCTTCAGCTCGCTGTCCCGAAGGTGAGTACATCGCCAGGGCCGGGTGGACCCGGGAT 1090
Db 1387 CCCTTCAGCTCGCTGTCCCGAAGGTGAGTACATCGCCAGGGCCGGGTGGACCCGGGAT 1446
QY 1091 GGCAGATACGCTGGGCACTGTCCTTGACCGGCCAGAGAGTGGCTCCAGCTGTCTC 1150
Db 1447 GGCAGATACGCTGGGCACTGTCCTTGACCGGCCAGAGAGTGGCTCCAGCTGTCTC 1506
QY 1151 CTCCTCCCGGCGCTGTTCATCCGAGCACAGAAATGAGAGCAGCGGCTAGCCCTTGCC 1210
Db 1507 CTCCTCCCGGCGCTGTTCATCCGAGCACAGAAATGAGAGCAGCGGCTAGCCCTTGCC 1566
QY 1211 AAGAGTGTCCCGAGAAATGTCACGCGTATGTGTGACAGAGAGTGCACCAACGCTGCG 1270
Db 1567 AAGAGTGTCCCGAGAAATGTCACGCGTATGTGTGACAGAGAGTGCACCAACGCTGCG 1626
QY 1271 ATCAATGTTCATACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGAGCTGTGC 1330
Db 1627 ATCAATGTTCATACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGAGAGCTGTGC 1686
QY 1331 TTCTCTCCGCGCAATGAATGCAAGACCGGCTTCTGCCATTTGTACAAAGTCAACGCCGTT 1390
Db 1687 TTCTCTCCGCGCAATGAATGCAAGACCGGCTTCTGCCATTTGTACAAAGTCAACGCCGTT 1746
QY 1391 TTAAATCCCAAGGCTACGATTTGGAGTGAACCTTCAGCCCGGGGAAGATGAATTTAAG 1450
|||||

Db 1747 TTAAATCCCAAGGCTACGATTTGGAGTGAACCTTCAGCCCGGGGAAGATGAATTTAAG 1806
QY 1451 TGCCCATTAAGGAAGAGATTTGCTGTACAGCGGTGAATGGAGGTTTGTGGAGGAC 1510
Db 1807 TGCCCATTAAGGAAGAGATTTGCTGTACAGCGGTGAATGGAGGTTTGTGGAGGAC 1866
QY 1511 GGCTCCAAATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTTCAGAGGACCAAGGAC 1570
Db 1867 GGCTCCAAATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTTCAGAGGACCAAGGAC 1926
QY 1571 ACGCGCTGAGAGACCAACCTCTACGTGTGATGAGAGCGCGCGGAGATCTGACG 1630
Db 1927 ACGCGCTGAGAGACCAACCTCTACGTGTGATGAGAGCGCGCGGAGATCTGACG 1986
QY 1631 CTCACCAAGCCCGGCTTCTCCATAGCTGTCCATGAGCCAGAACTTCGACATGTTGTC 1890
Db 1987 CTCACCAAGCCCGGCTTCTCCATAGCTGTCCATGAGCCAGAACTTCGACATGTTGTC 2046
QY 1691 AGCCACTACAGACGCTGAGACGCGCCCTGGGTGCACGCTTACAAAGTGAAGGCGCC 1750
Db 2047 AGCCACTACAGACGCTGAGACGCGCCCTGGGTGCACGCTTACAAAGTGAAGGCGCC 2106
QY 1751 GACGAGACCCCTGCAACAGACGCCCGCTTGGGCTAGCATGATGAGACGCCAGC 1810
Db 2107 GACGAGACCCCTGCAACAGACGCCCGCTTGGGCTAGCATGATGAGACGCCAGC 2166
QY 1811 TGCCCCCGGATTAATGTTCTCCAGAGATCTTCATTTCCACAGCGGCTGATGTCGG 1870
Db 2167 TGCCCCCGGATTAATGTTCTCCAGAGATCTTCATTTCCACAGCGGCTGATGTCGG 2226
QY 1871 CTCTAGGCGATGATCTAACAGCCCAAGCCCTTGACGCCAGAGGAAGACCCACCGTC 1930
Db 2227 CTCTAGGCGATGATCTAACAGCCCAAGCCCTTGACGCCAGAGGAAGACCCACCGTC 2286
QY 1931 CTCTTTGATATGAGAGGCCCCAGGTGCACTGAGTGAATACCTTCCTTCAAGGATCAAG 1990
Db 2287 CTCTTTGATATGAGAGGCCCCAGGTGCACTGAGTGAATACCTTCCTTCAAGGATCAAG 2346
QY 1991 TACTTGCGGCTCAACACACTGAGCCTCCCTGGGCTACGCGCTGTGATGAGGCGAG 2050
Db 2347 TACTTGCGGCTCAACACACTGAGCCTCCCTGGGCTACGCGCTGTGATGAGGCGAG 2406
QY 2051 GGCTCTGTGACGAGGGCTTGGGTTGAAAGGGCCCTGAACCAATGGGCCAGGTG 2110
Db 2407 GGCTCTGTGACGAGGGCTTGGGTTGAAAGGGCCCTGAACCAATGGGCCAGGTG 2466
QY 2111 GAGATGAGAGCAGGTGAGGGGCTGCAATGCTGGCCGAGAAAGTATGCTCATGCGAC 2170
Db 2467 GAGATGAGAGCAGGTGAGGGGCTGCAATGCTGGCCGAGAAAGTATGCTCATGCGAC 2526
QY 2171 CTGAGCGGAGTTCATTCATCTGCTGTGCTTACGAGGGGCTTCTCTGCTCATGAGGCTA 2230
Db 2527 CTGAGCGGAGTTCATTCATCTGCTGTGCTTACGAGGGGCTTCTCTGCTCATGAGGCTA 2586
QY 2231 ATCCACAAGCCCCAGGTGTTCAAGGTGGCCATGCGGAGTGCACCGTCTGGATG 2290
Db 2587 ATCCACAAGCCCCAGGTGTTCAAGGTGGCCATGCGGAGTGCACCGTCTGGATG 2646
QY 2291 GGCTTACGACACAGGGTACACTGAGGGCTACATGAGAGTCCCTGAGAAACAACGACGGC 2350
Db 2647 GGCTTACGACACAGGGTACACTGAGGGCTACATGAGAGTCCCTGAGAAACAACGACGGC 2706
QY 2351 TATGAGGCGGGTTCCGTGGCCCTGACAGTGGAGAACTGGCCCAATGAGCCCAAGCCTTG 2410
Db 2707 TATGAGGCGGGTTCCGTGGCCCTGACAGTGGAGAACTGGCCCAATGAGCCCAAGCCTTG 2766
QY 2411 CTATATCTCCAGGGCTTCTGACAGAAACGTGACATTTTTCACACAACTTCTCTGTC 2470
Db 2767 CTATATCTCCAGGGCTTCTGACAGAAACGTGACATTTTTCACACAACTTCTCTGTC 2826
QY 2471 TCCCACTGATCCGACAGAGGAACCTTACAGCTCCCAATCTACCCCAACGAGAGAC 2530
Db 2827 TCCCACTGATCCGACAGAGGAACCTTACAGCTCCCAATCTACCCCAACGAGAGAC 2886
|||||

QY 2531 AGATTGCGTGGCCGAGTGGGCGAGACATATGAATGCTTGGTGCACCTTTCTACAG 2590
|||||
Db 2887 AGATTGCGTGGCCGAGTGGGCGAGACATATGAATGCTTGGTGCACCTTTCTACAG 2946
QY 2591 GAATACCTCTGAGC 2604
|||||
Db 2947 GAATACCTCTGAGC 2960
RESULT 3
ABK83333
ID ABK83333 standard; cDNA; 4302 BP.
XX
AC ABK83333;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPRP-2 splice variant #1.
XX
KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyskinesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001MO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
XX
P-PSDB; ABG61602.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
PS Disclosure; Page 78-79; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ Sequence 4302 BP; 923 A; 1350 C; 1221 G; 808 T; 0 other;

Query Match 99.1%; Score 2592.4; DB 24; Length 4302;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2593; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 ATGGCCACACCGGAGACCCCAAGCGCCAGCGAGCGACCGCCACAGATGACCG 70
|||||
Db 367 ATGGCCACACCGGAGACCCCAAGCGCCAGCGAGCGACCGCCACAGATGACCG 426
QY 71 GCCGCCGCTTCCAGGTGCAAGAAAGCACTCGTGGAGCGGCTCCGAGCATATCAACGGC 130
|||||
Db 427 GCCGCCGCTTCCAGGTGCAAGAAAGCACTCGTGGAGCGGCTCCGAGCATATCAACGGC 486
QY 131 AGCCGCAAGTACTGCGGCGCTCATTTGTCACAAAGCGCCCGACACTTCCAGTTTGTCAG 190
|||||
Db 487 AGCCGCAAGTACTGCGGCGCTCATTTGTCACAAAGCGCCCGACACTTCCAGTTTGTCAG 546
QY 191 AAGACGATGATGCTGAGCGCCCGACCTCCACCGGCTCTACTACTGAGGAATGCCATATGCG 250
|||||
Db 547 AAGACGATGATGCTGAGCGCCCGACCTCCACCGGCTCTACTACTGAGGAATGCCATATGCG 606
QY 251 AGCCGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAAAGTCCGGAAGAGGCTCTG 310
|||||
Db 607 AGCCGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAAAGTCCGGAAGAGGCTCTG 666
QY 311 CTGCTCTGCTCTGCGAAGCAGATGCTGATCATTTCCAGGCGCACCGCCACATGGGCTC 370
|||||
Db 667 CTGCTCTGCTCTGCGAAGCAGATGCTGATCATTTCCAGGCGCACCGCCACATGGGCTC 726
QY 371 TACTCTCGGAGGAGGAGAGCTGCTGAGGAGCGGAAAGCGCTGGGGGCTTCGGCATCACG 430
|||||
Db 727 TACTCTCGGAGGAGGAGAGCTGCTGAGGAGCGGAAAGCGCTGGGGGCTTCGGCATCACG 786
QY 431 TCCTAGCACTTCCACAGCGAGAGTGGGCTTCCTCTCCAGGCGCACAGACCTCTTC 490
|||||
Db 787 TCCTAGCACTTCCACAGCGAGAGTGGGCTTCCTCTCCAGGCGCACAGACCTCTTC 846
QY 491 CACTGTGCGAGCGCGCGCAAGAAAGCGCTTATGCTGTCCCTTATGAACCGCTGGAATTC 550
|||||
Db 847 CACTGTGCGAGCGCGCGCAAGAAAGCGCTTATGCTGTCCCTTATGAACCGCTGGAATTC 906
QY 551 AAGACCGAGTGTCAAGGGCCCGGATGAGACCCCAAAATCGCCCTGCGACCGCTTCG 610
|||||
Db 907 AAGACCGAGTGTCAAGGGCCCGGATGAGACCCCAAAATCGCCCTGCGACCGCTTCG 966
QY 611 TTCTCTTCATCAATACAGACGACCTGTGGTGGCCAAATCGACAGAGCGAGAGCGG 670
|||||
Db 967 TTCTCTTCATCAATACAGACGACCTGTGGTGGCCAAATCGACAGAGCGAGAGCGG 1026
QY 671 CGGCTGACTTTCGCCACCAAGGTTTATCCAAATGCTCGATGACCCCAAGTCCGGGT 730
|||||
Db 1027 CGGCTGACTTTCGCCACCAAGGTTTATCCAAATGCTCGATGACCCCAAGTCCGGGT 1086
QY 731 GTGGCCACCTTCGTATACAGGAAGTGTGACCGGCTCACTGGGTACTGGTGTGCCCC 790
|||||
Db 1087 GTGGCCACCTTCGTATACAGGAAGTGTGACCGGCTCACTGGGTACTGGTGTGCCCC 1146
QY 791 ACAGCCCTCTGGGAAGGTTTCAGAGGCGCTCAAGCGCTCGGAATCTGTATGAGGAATG 850
|||||
Db 1147 ACAGCCCTCTGGGAAGGTTTCAGAGGCGCTCAAGCGCTCGGAATCTGTATGAGGAATG 1206
QY 851 GATGAGTCCGAGGTGAGGTGATTCACGTCCCTCTCTCGCTAGAGAAAGGAAGACG 910
|||||
Db 1207 GATGAGTCCGAGGTGAGGTGATTCACGTCCCTCTCTCGCTAGAGAAAGGAAGACG 1266
QY 911 GATCGGATGAGGATACCCGAGGACAGGACGAGCAAGAAATGCCCTTGAAGTGGCT 970
|||||
Db 1267 GATCGGATGAGGATACCCGAGGACAGGACGAGCAAGAAATGCCCTTGAAGTGGCT 1326
QY 971 GAGTTCAGAGTACAGACCGAGGCAAGATGCTCGACCCAGGAGAAAGAGAGTGTGTCAG 1030
|||||
Db 1327 GAGTTCAGAGTACAGACCGAGGCAAGATGCTCGACCCAGGAGAAAGAGAGTGTGTCAG 1386
QY 1031 CCCTTCAGTCTGCTGTTCCGGAAGGTGAGTACATGCGCAGGGCCGGGTGACCCGGAT 1090
|||||
Db 1387 CCCTTCAGTCTGCTGTTCCGGAAGGTGAGTACATGCGCAGGGCCGGGTGACCCGGAT 1446
QY 1091 GGCAAAATACGCTGGGCGCATGTTCTTGAGACCGGCCACAGATGCGTCCAGCTGCTCTC 1150

```

|||||
Db 1447 GCAAAATACGCTGGGCGCATGTTCTCGAGCCGCGCCACAGCTGGCTCCAGCTCGCTC 1506
Oy 1151 CTCGCCCCGGGCTGTATATCCGAGACAGAGAAATGAGAGACAGGGGTACTGCTGTC 1210
Db 1507 CTCGCCCCGGGCTGTATATCCGAGACAGAGAAATGAGAGACAGGGGTACTGCTGTC 1566
Oy 1211 AGAGCTGCCAGGAGATGTCCAGCGGTATGTGTGTACGAGAGGTCACCAAGCTGTGG 1270
Db 1567 AGAGCTGCCAGGAGATGTCCAGCGGTATGTGTGTGTACGAGAGGTCACCAAGCTGTGG 1626
Oy 1271 ATCAATGTTTCATGACATCTTCTATCCCTTCCCAATCAGAGAGAGAGACAGCTCTGC 1330
Db 1627 ATCAATGTTTCATGACATCTTCTATCCCTTCCCAATCAGAGAGAGAGACAGCTCTGC 1686
Oy 1331 TTTTCGCGGCGCAATGAATGCAAGACCGGCTCTGCGCATTTGTACAAATGTCACCGCGTT 1390
Db 1687 TTTTCGCGGCGCAATGAATGCAAGACCGGCTCTGCGCATTTGTACAAATGTCACCGCGTT 1746
Oy 1391 TTTAAATCCAGGGCTACGATTGAGGTGAGCCCTTCAGCCCGGGGAAAGATGAATTTAAG 1450
Db 1747 TTTAAATCCAGGGCTACGATTGAGGTGAGCCCTTCAGCCCGGGGAAAGATGAATTTAAG 1806
Oy 1451 TCCGCGCATTAAGAGAGAGATTGCTGTGACACGCGTGAATGGAGGTTTGGCGAGGCAC 1510
Db 1807 TGCCCGCATTAAGAGAGAGATTGCTGTGACACGCGTGAATGGAGGTTTGGCGAGGCAC 1866
Oy 1511 GGCTCCAGAGATGTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGCGCAAGAGAGAC 1570
Db 1867 GGCTCCAGAGATGTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGCGCAAGAGAGAC 1926
Oy 1571 AGCGCGCTGAGACACACCTCTACGTGTGTCAGCTATGAGGCGCGCGAGATCTGACGC 1630
Db 1927 AGCGCGCTGAGACACACCTCTACGTGTGTCAGCTATGAGGCGCGCGAGATCTGACGC 1986
Oy 1631 CTCACACAGCGCGGCTTCTCCCATAGCTGCTCATGAGCCAGAACTTGCACATGTTCTGC 1690
Db 1987 CTCACACAGCGCGGCTTCTCCCATAGCTGCTCATGAGCCAGAACTTGCACATGTTCTGC 2046
Oy 1691 AGCGACTACAGAGGCTGAGAGACGCGCGCTGCGTGCAGCTGTACAAAGTGTAGCGGCGCC 1750
Db 2047 AGCGACTACAGAGGCTGAGAGACGCGCGCTGCGTGCAGCTGTACAAAGTGTAGCGGCGCC 2106
Oy 1751 GACGACGAGCCCCCTGCACAGAGCCCGCTTCTGCGGTAGCATGATGAGAGGACGACGAC 1810
Db 2107 GACGACGAGCCCCCTGCACAGAGCCCGCTTCTGCGGTAGCATGATGAGAGGACGACGAC 2166
Oy 1811 TGCCCCCGGAGATTATGTTCTCCAGAGATCTTCCATTTCCACACGCGCTGCGATGTGCGG 1870
Db 2167 TGCCCCCGGAGATTATGTTCTCCAGAGATCTTCCATTTCCACACGCGCTGCGATGTGCGG 2226
Oy 1871 CTCACGCGCATGTCTACAAAGCCCGCACGCGCTTGCAGCCAGGAGAGAGACCCACCGCTC 1930
Db 2227 CTCACGCGCATGTCTACAAAGCCCGCACGCGCTTGCAGCCAGGAGAGAGACCCACCGCTC 2286
Oy 1931 CTCTTTGTATATGAGGCCCCAGGATGACGCTGTGAATTAACCTTCTCAAAAGCATGAAG 1990
Db 2287 CTCTTTGTATATGAGGCCCCAGGATGACGCTGTGAATTAACCTTCTCAAAAGCATGAAG 2346
Oy 1991 TACTTGGCGGCTCAACACACTGGCTCTCCCTGAGCTACGCGCTGTGTGTATGACGCGCAGG 2050
Db 2347 TACTTGGCGGCTCAACACACTGGCTCTCCCTGAGCTACGCGCTGTGTGTATGACGCGCAGG 2406
Oy 2051 GGCTCCTGTCAAGAGGCGTGGTTTGAAGGCGCTGAAGAACCAAAATGGGCGCAGGTTG 2110
Db 2407 GGCTCCTGTCAAGAGGCGTGGTTTGAAGGCGCTGAAGAACCAAAATGGGCGCAGGTTG 2466
Oy 2111 GAGATCGAGAGACAGAGTGGAGGCTGTGAGTTCTGTGCGCGAGAGATGATGGCTTCAATGAC 2170
Db 2467 GAGATCGAGAGACAGAGTGGAGGCTGTGAGTTCTGTGCGCGAGAGATGATGGCTTCAATGAC 2526
Oy 2171 CTGAGCCGAGTTGCCATCATGCTGTGTCTTACGGGGGCTTCTCTCGCTCATGGGCGCTA 2230
|||||

```

```

Db 2527 CTGAGCCGAGTTGCCATCCAGTGGCTGTCTTACGGGGGCTTCTCTGCTCATGGGCTTA 2586
Oy 2231 ATCCACAAGCCCCAGAGTGTTCAAAGTGGCCATGCGGGTGCCTCCGGTACACGCTTGATG 2290
Db 2587 ATCCACAAGCCCCAGAGTGTTCAAAGTGGCCATGCGGGTGCCTCCGGTACACGCTTGATG 2646
Oy 2291 GCGTCAGACAGAGGTTACACTGAGCGCTACATGACGTCCTGTGAGACAAACAGCAGCGGC 2350
Db 2647 GCGTCAGACAGAGGTTACACTGAGCGCTACATGACGTCCTGTGAGACAAACAGCAGCGGC 2706
Oy 2351 TATGAGGGGGGTTCCGTGGCCCTGACGTGGAGAGCTGCCCCAATGAGCCCAACCGCTTG 2410
Db 2707 TATGAGGGGGGTTCCGTGGCCCTGACGTGGAGAGCTGCCCCAATGAGCCCAACCGCTTG 2766
Oy 2411 CTATCTCTCACAGGCTTCTCTGAGCAAAAGCTGCACTTTTTCACACAACCTTCTCTGTC 2470
Db 2767 CTATCTCTCACAGGCTTCTCTGAGCAAAAGCTGCACTTTTTCACACAACCTTCTCTGTC 2826
Oy 2471 TCCCAACTGATCCGAGACAGAGGAAACCTTACACAGCTTCACAGATACCCCAAGAGAGAC 2530
Db 2827 TCCCAACTGATCCGAGACAGAGGAAACCTTACACAGCTTCACAGATACCCCAAGAGAGAC 2886
Oy 2531 AGTATTCGTCGCGCGAGTGGCGGAGACACTATGAAGTCAGCTGTGCTGACTTCTTACAG 2590
Db 2887 AGTATTCGTCGCGCGAGTGGCGGAGACACTATGAAGTCAGCTGTGCTGACTTCTTACAG 2946
Oy 2591 GAATACCTCTGAGC 2604
Db 2947 GAATACCTCTGAGC 2960

```

```

RESULT 4
AAD38954
ID AAD38954 standard; cDNA; 3024 BP.
XX
AC AAD38954;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 9 (DPP9) cDNA.
XX
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immunodeficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2910
FT /tag= a
FT /product= "Human DPP9 protein"
FT /transl_except= (pos: 1120..1122, aa:Glu)
FT /note= "CDS does not include start codon"
FT /partial
XX
MO200234900-A1.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-AU01388.
XX
PR 27-OCT-2000; 2000AU-0001078.
XX
PA (UNSY ) UNIV SYDNEY.
XX
PI Abbott CA, Gorrell MD;
XX
DR WPI: 2002-454646/48.
XX
DR P-PSDB: AAE24168.
XX
PT New dipeptidyl peptidase (DPP) peptides, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat

```

PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
rejection and HIV infection -
XX
PS Example; Fig 4; 91pp; English.

CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP9 cDNA.

XX Sequence 3024 BP; 624 A; 973 C; 875 G; 552 T; 0 other;

Query Match 98.9%; Score 2587.6; DB 24; Length 3024;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2550; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 11 ATGGCCACACCGGGAGCCCGACGAGCGAGCCAGCGCCGACAGATGACCG 70
DB 319 ATGGCCACACCGGGAGCCCGACGAGCGAGCGAGCGCCGACAGATGACCG 378
QY 71 GCGCCCGCTTCCAGGTGCGAAGCACTGCTGGAGCGGCTCGGAGCATTCACGCG 130
DB 379 GCGCCCGCTTCCAGGTGCGAAGCACTGCTGGAGCGGCTCGGAGCATTCACGCG 438
QY 131 AGCGGCAAGTCTGGGCGCTCATTTGCAAGAAGCGCCCGACACTTCCAGTTTGTCCAG 190
DB 439 AGCGGCAAGTCTGGGCGCTCATTTGCAAGAAGCGCCCGACACTTCCAGTTTGTCCAG 498
QY 191 AAGAGGATGATCTGGGCGCCCACTCCACCGCTTACTACCTGGGAATGCCATATGGC 250
DB 499 AAGAGGATGATCTGGGCGCCCACTCCACCGCTTACTACCTGGGAATGCCATATGGC 558
QY 251 AGCGGAGAGATCTCTCTCTACTCTGAGATTCACAGAAGCTCGGAAAGAGGCTCTG 310
DB 559 AGCGGAGAGATCTCTCTCTACTCTGAGATTCACAGAAGCTCGGAAAGAGGCTCTG 618
QY 311 CTGCTCTCTCTGGAAGAGATGCTGATCATTTCCAGGCGACGGCCCAACGATGGGCTC 370
DB 619 CTGCTCTCTCTGGAAGAGATGCTGATCATTTCCAGGCGACGGCCCAACGATGGGCTC 678
QY 371 TACTCTCGGAGAGAGAGTCTGAGGAGCGGAAAGCGCTGGGGGCTTTCGGGATACC 430
DB 679 TACTCTCGGAGAGAGAGTCTGAGGAGCGGAAAGCGCTGGGGGCTTTCGGGATACC 738
QY 431 TCTAGAGACTTCCACAGCGAGAGTGGCTTCTCTTCCAGGCGACAAAGCTCTTTC 490
DB 739 TCTAGAGACTTCCACAGCGAGAGTGGCTTCTCTTCCAGGCGACAAAGCTCTTTC 798
QY 491 CACTGTGCGAGCGGCGCAAGAGGCTTATGATGTGCCCTATGAACCGCTGGGAATC 550
DB 799 CACTGTGCGAGCGGCGCAAGAGGCTTATGATGTGCCCTATGAACCGCTGGGAATC 858
QY 551 AAGACCCAGTCTCAGGGGCGCGGATGAGCCCAAAATCTGCCTGCGAGCCCTGCTTC 610
DB 859 AAGACCCAGTCTCAGGGGCGCGGATGAGCCCAAAATCTGCCTGCGAGCCCTGCTTC 918
QY 611 TTCTCTCTTCAATAACAGCGACCTGTGGTGGCCAACTCGAGACAGCGAGAGCGG 670
DB 919 TTCTCTCTTCAATAACAGCGACCTGTGGTGGCCAACTCGAGACAGCGAGAGCGG 978
QY 671 CGGCTGACCTTGTGCCACCAAGGTTTATCAATGTCTGTGATGACCCCAAGTCTGGGGT 730
DB 979 CGGCTGACCTTGTGCCACCAAGGTTTATCAATGTCTGTGATGACCCCAAGTCTGGGGT 1038
QY 731 GTGGGCAACCTTCTCATACAGAAAGTTCGACCGCTTCACTGGGTACTGGTGGTCCC 790
DB 1039 GTGGGCAACCTTCTCATACAGAAAGTTCGACCGCTTCACTGGGTACTGGTGGTCCC 1098
QY 791 ACAGCTCTCTGGGAAGGTTAGAGGGCTTAAGACGCTGGGAATCTGTATGAGGAATC 850
DB 1099 ACAGCTCTCTGGGAAGGTTAGAGGGCTTAAGACGCTGGGAATCTGTATGAGGAATC 1158

QY 851 GATGAGTCCGAGGAGGAGTTCATTCACGTCCCTCTCGGCTAGAGAAGAAAGAACAGC 910
DB 1159 GATGAGTCCGAGGAGGAGTTCATTCACGTCCCTCTCTCGGCTAGAGAAGAAAGAACAGC 1218
QY 911 GACTCGTATGAGTACCCAGAGAGAGGAGAGCAAGATCCCAAGATTGGCTTGAACCTGGCT 970
DB 1219 GACTCGTATGAGTACCCAGAGAGAGGAGAGCAAGATCCCAAGATTGGCTTGAACCTGGCT 1278
QY 971 GAGTTCAGACTGACAGCCAGGGCAGAGATGCTTCGACCCAGAGAGAAGAGCTGTGTCCAG 1030
DB 1279 GAGTTCAGACTGACAGCCAGGGCAGAGATGCTTCGACCCAGAGAGAAGAGCTGTGTCCAG 1338
QY 1031 CCCCTGAGCTGGCTTCCCGGAAGGTGAGTACATCCCGAGGGCGGGGTGAGCCGGGAT 1090
DB 1339 CCCCTGAGCTGGCTTCCCGGAAGGTGAGTACATCCCGAGGGCGGGGTGAGCCGGGAT 1398
QY 1091 GGCMAATACGCTTGGGCGCATTTCTTGAGCCGGCCAGAGAGTGGCTCCAGCTGTCTCC 1150
DB 1399 GGCMAATACGCTTGGGCGCATTTCTTGAGCCGGCCAGAGAGTGGCTCCAGCTGTCTCC 1458
QY 1151 CTCCCGCCGCGCTTTCATCCGAGCAGAGAGATGAGAGCAGCGGCTAGCCTGTGCC 1210
DB 1459 CTCCCGCCGCGCTTTCATCCGAGCAGAGAGATGAGAGCAGCGGCTAGCCTGTGCC 1518
QY 1211 AGAGCTGTCCCGAGAGATGTCCAGCCGTATGTGTGACAGAGAGTCCACCAAGCTGTG 1270
DB 1519 AGAGCTGTCCCGAGAGATGTCCAGCCGTATGTGTGACAGAGAGTCCACCAAGCTGTG 1578
QY 1271 ATCAATGTTATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGGAGAGAGCTGTGC 1390
DB 1579 ATCAATGTTATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGGAGAGAGCTGTGC 1638
QY 1331 TTTCTCCGCGCAATGAGTSCAAGACCGGCTTCTGCCATTTGTACAAGTCCACCGCGTT 1390
DB 1639 TTTCTCCGCGCAATGAGTSCAAGACCGGCTTCTGCCATTTGTACAAGTCCACCGCGTT 1698
QY 1391 TTTAAATCCAGGCGCTACGATTTGGAGTACCCCTTCAGCCCGCGGGAAGATGAATTTAAG 1450
DB 1699 TTTAAATCCAGGCGCTACGATTTGGAGTACCCCTTCAGCCCGCGGGAAGATGAATTTAAG 1758
QY 1451 TGGCCCATTTAAGAGAGATGTGCTGTGACCAAGCGGTAAATGGAGGTTTGGGAGGAC 1510
DB 1759 TGGCCCATTTAAGAGAGATGTGCTGTGACCAAGCGGTAAATGGAGGTTTGGGAGGAC 1818
QY 1511 GGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCACCAAGGAC 1570
DB 1819 GGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCACCAAGGAC 1878
QY 1571 AGCGCGCTGAGAGACCAACGCTACGTGTGACGATGAGAGCGCGCGGAGATCGTACGC 1630
DB 1879 AGCGCGCTGAGAGACCAACGCTACGTGTGACGATGAGAGCGCGCGGAGATCGTACGC 1938
QY 1631 CTCACACAGCGCGGCTTCTCCATAGCTGTGCTCATAGGAGCGAGAACTTCGACATGTTGCTC 1690
DB 1939 CTCACACAGCGCGGCTTCTCCATAGCTGTGCTCATAGGAGCGAGAACTTCGACATGTTGCTC 1998
QY 1691 AGCCACTACAGCAGCGTGAAGCGCGGCTGTGCTGACGTCTTACAAAGCTGAGCGGCCCC 1750
DB 1999 AGCCACTACAGCAGCGTGAAGCGCGGCTGTGCTGACGTCTTACAAAGCTGAGCGGCCCC 2058
QY 1751 GAGGAGAGCGCGCTGCAAGAGAGCGCGGCTGTGGGCTTACGATGAGGAGGAGCGGACG 1810
DB 2059 GAGGAGAGCGCGCTGCAAGAGAGCGCGGCTGTGGGCTTACGATGAGGAGGAGCGGACG 2118
QY 1811 TGGCCCGCGGATTTATGTTCTCCAGAGATCTTCCATTTCCACAGCGCTGAGATGTGCGG 1870
DB 2119 TGGCCCGCGGATTTATGTTCTCCAGAGATCTTCCATTTCCACAGCGCTGAGATGTGCGG 2178
QY 1871 CTCTAGGCGATGATCTTACAAGCGCCCTTGACGCGAGGAGAAAGACCCACCGCTC 1930
DB 2179 CTCTAGGCGATGATCTTACAAGCGCCCTTGACGCGAGGAGAAAGACCCACCGCTC 2238

QY 1931 CTCTTTGATATGAGAGCCCGGAGGTGAGTGTGATTAATCTCTTCAAAAGCATCAAG 1990
 |||||
 Db 2239 CTCTTTGATATGAGAGCCCGGAGGTGAGTGTGATTAATCTCTTCAAAAGCATCAAG 2298
 QY 1991 TACTTGGGCTCAACACACTGGGCTCCCTGAGGTAGCCGCTGTTGATGATGAGCAGG 2050
 |||||
 Db 2299 TACTTGGGCTCAACACACTGGGCTCCCTGAGGTAGCCGCTGTTGATGATGAGCAGG 2358
 QY 2051 GGTCTCTGTCAGAGAGGCTTCTGTTGAAAGGGGCTGAAAAACCAATGGGCGAGGTG 2110
 |||||
 Db 2359 GGTCTCTGTCAGAGAGGCTTCTGTTGAAAGGGGCTGAAAAACCAATGGGCGAGGTG 2418
 QY 2111 GAGATCGAGAGCAGGTGAGAGGCTGTCAGATTGCTGAGCCAGAGATGATGCTTATCGAC 2170
 |||||
 Db 2419 GAGATCGAGAGCAGGTGAGAGGCTGTCAGATTGCTGAGCCAGAGATGATGCTTATCGAC 2478
 QY 2171 CTGAGCCGAGTTGCCATTCATGGCTGTCTAGGGGGCTTCTCTGCTCATAGGGGCTA 2230
 |||||
 Db 2479 CTGAGCCGAGTTGCCATTCATGGCTGTCTAGGGGGCTTCTCTGCTCATAGGGGCTA 2538
 QY 2231 ATCCACAAGCCCGAGGTGTTCAAGTGGCCATGCGGGGTGCCCGGTCAACGCTGGATG 2290
 |||||
 Db 2539 ATCCACAAGCCCGAGGTGTTCAAGTGGCCATGCGGGGTGCCCGGTCAACGCTGGATG 2598
 QY 2291 GCCTACGACACAGGCTACACTGAGCCCTACATGAGCGTCCCTAGAGAACACACAGCGGC 2350
 |||||
 Db 2599 GCCTACGACACAGGCTACACTGAGCCCTACATGAGCGTCCCTAGAGAACACACAGCGGC 2658
 QY 2351 TATGAGGCGGGTTCGTCGGGCTGTCAGGTGAGAGGTGCCCAATGAGCCCAACCGCTTG 2410
 |||||
 Db 2659 TATGAGGCGGGTTCGTCGGGCTGTCAGGTGAGAGGTGCCCAATGAGCCCAACCGCTTG 2718
 QY 2411 CTATATCCTCCACGCTTCTCTGAGAGAAAGTGCACATTTTCCACACAACCTTCTCTGTC 2470
 |||||
 Db 2719 CTATATCCTCCACGCTTCTCTGAGAGAAAGTGCACATTTTCCACACAACCTTCTCTGTC 2778
 QY 2471 TCCCAACTGATCCGAGCAGGAGAAACCTTACCACTCCAGTCTTACCCCAAGAGACAC 2530
 |||||
 Db 2779 TCCCAACTGATCCGAGCAGGAGAAACCTTACCACTCCAGTCTTACCCCAAGAGACAC 2838
 QY 2531 AGTATTCGCTGGCCGAGTGGGGGAGCACTATGAAAGTCACTTGGTGCACCTTCTACAG 2590
 |||||
 Db 2839 AGTATTCGCTGGCCGAGTGGGGGAGCACTATGAAAGTCACTTGGTGCACCTTCTACAG 2898
 QY 2591 GAATACCTCTGAGC 2604
 |||||
 Db 2899 GAATACCTCTGAGC 2912
 RESULT 5
 ABR83339
 ID ABR83339 standard; cDNA: 4180 BP.
 XX
 AC ABR83339;
 DE
 DT 12-AUG-2002 (first entry)
 XX
 DE cDNA encoding human DPRP-2 splice variant #7.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychiatric disorder; neurological disorder;
 KW dyslexia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 OS Homo sapiens.
 PN MO200231134-A2.
 PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 PI Qi S, Akinsanya KO, Riviere PJ, Junlien J;
 PI
 DR MPI; 2002-444178/47.
 DR P-PSDB; ABG61608.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 PS Disclosure; Page 97-98; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyslexias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABR83322-ABR83343 encode human DPRP proteins.
 XX
 SQ Sequence 4180 BP; 898 A; 1312 C; 1178 G; 792 T; 0 other;
 Query Match 95.7%; Score 2504.4; DB 24; Length 4180;
 Best local Similarity 98.5%; Pred. No. 0;
 Matches 2554; Conservative 0; Mismatches 1; Indels 39; Gaps 1;
 QY 11 ATGGCCACACCGGGAGCCCAACGCGCGAGCGAGGAGCAGCCGCCACAGATGACCGG 70
 |||||
 Db 367 ATGGCCACACCGGGAGCCCAACGCGCGAGCGAGGAGCAGCCGCCACAGATGACCGG 426
 QY 71 GCCGCCGCTTCCAGGTGAGAGACACTGCTGGAGCGGCTCCGAGCATCATCCAGCGC 130
 |||||
 Db 427 GCCGCCGCTTCCAGGTGAGAGACACTGCTGGAGCGGCTCCGAGCATCATCCAGCGC 486
 QY 131 AGCCGCAAGTACTCGGGCTCATTTGCAACAGAGCGCCCGCAGCATTTCCAGTTGGCAG 190
 |||||
 Db 487 AGCCGCAAGTACTCGGGCTCATTTGCAACAGAGCGCCCGCAGCATTTCCAGTTGGCAG 546
 QY 191 AAGAGGATGATCTGGGCGCCCACTCCACCGGCTTACTACTCTGGGAATGCAATGAGC 250
 |||||
 Db 547 AAGAGGATGATCTGGGCGCCCACTCCACCGGCTTACTACTCTGGGAATGCAATGAGC 606
 QY 251 AGCCGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAGGTCCGGAAGAGGCTTG 310
 |||||
 Db 607 AGCCGAGAGAACTCCCTCTCTCTACTCTGAGATTCCCAAGAGGTCCGGAAGAGGCTTG 666
 QY 311 CTGCTCTGCTCTGGAAGAGATGCTGATCATTTCCAGGCCACGCCCAACATGGGGTG 370
 |||||
 Db 667 CTGCTCTGCTCTGGAAGAGATGCTGATCATTTCCAGGCCACGCCCAACATGGGGTG 726
 QY 371 TACTTCGGAGAGAGAGTGTGAGGAGCGGAACGCTGGGGTCTTCCGATCACAC 430
 |||||
 Db 727 TACTTCGGAGAGAGAGTGTGAGGAGCGGAACGCTGGGGTCTTCCGATCACAC 786
 QY 431 TCTTACGACTTCCACAGCAGAGTGGCTCTTCTCTTCCAGGCCACAGAACAGCCTCTTC 490
 |||||
 Db 787 TCTTACGACTTCCACAGCAGAGTGGCTCTTCTCTTCCAGGCCACAGAACAGCCTCTTC 846
 QY 491 CACTGTGCGAGCGCGGCAAGAAAGGCTTTCANGTGTCCCTATGAAACCGCTGAAATC 550
 |||||

Db 847 CACTGCCGAGGCGGCAAGAACGGCTTCATGTGTCCCTATGAAACCGCTGAAATC 906
 QY 551 AAGACCCAGTGCATGAGGCGCCGGATGAGACCCAAATGTGCCCTGACCCCTGCTTC 610
 Db 907 AAGACCCAGTGCATGAGGCGCCGGATGAGACCCAAATGTGCCCTGACCCCTGCTTC 966
 QY 611 TTCTCTTCATCAATTAACAGACGACCTGTGGTGCCAAATGAGACGAGGCGAGCGG 670
 Db 967 TTCTCTTCATCAATTAACAGACGACCTGTGGTGCCAAATGAGACGAGGCGAGCGG 1026
 QY 671 CGGCTGACCTTCTGCCCAAGGTTTATCCAAATGTCTGTGATGACCCCAAGTCTGCGGT 730
 Db 1027 CGGCTGACCTTCTGCCCAAGGTTTATCCAAATGTCTGTGATGACCCCAAGTCTGCGGT 1086
 QY 731 GTGGCCACCTTGTCTATTAAGGAAGAGTTCGACCCGTTCACTGGGTACTGTGTGCTCC 790
 Db 1087 GTGGCCACCTTGTCTATTAAGGAAGAGTTCGACCCGTTCACTGGGTACTGTGTGCTCC 1146
 QY 791 ACAGCCCTCTGGGAAGGTTTCAGAGGCTTCAGAGCGCTGGAATCCTGTATGAGAAATC 850
 Db 1147 ACAGCCCTCTGGGAAGGTTTCAGAGGCTTCAGAGCGCTGGAATCCTGTATGAGAAATC 1206
 QY 851 GATGAGTCCGAGGTGAGGTCAATTCACGTCCCTCTCTGCGCTAGAGAAGAAAGAGAGC 910
 Db 1207 GATGAGTCCGAGGTGAGGTCAATTCACGTCCCTCTCTGCGCTAGAGAAGAAAGAGAGC 1266
 QY 911 GACTGCTATCGGTATACCCCAAGAGAGAGAGAGATCCCAAGATTCCTTGAATCTGCT 970
 Db 1267 GACTGCTATCGGTATACCCCAAGAGAGAGAGAGATCCCAAGATTCCTTGAATCTGCT 1326
 QY 971 GAGTTCAGACTGACAGAGGCGCAAGATCTGTCAGACGAGAGAAGAGAGAGTGTGAG 1030
 Db 1327 GAGTTCAGACTGACAGAGGCGCAAGATCTGTCAGACGAGAGAAGAGAGAGTGTGAG 1386
 QY 1031 CCCTTCAGCTCGCTGTTCCTCGAAGGTGAGATCATCCAGAGGCGCGGTGAGACCCGGAT 1090
 Db 1387 CCCTTCAGCTCGCTGTTCCTCGAAGGTGAGATCATCCAGAGGCGCGGTGAGACCCGGAT 1446
 QY 1091 GCGAATAGGCGTGGGCGCAATGTTCTGTGAGACCGGCGCCAGACAGTGGCTCCAGCTGCTCTC 1150
 Db 1447 GCGAATAGGCGTGGGCGCAATGTTCTGTGAGACCGGCGCCAGAGTGGCTCCAGCTGCTCTC 1506
 QY 1151 CTCGCCCGGCGCTTTCATCCCGAGACAGAGAAATGAGAGAGGCGGTAGCGCTTGCC 1210
 Db 1507 CTCGCCCGGCGCTTTCATCCCGAGACAGAGAAATGAGAGAGGCGGTAGCGCTTGCC 1566
 QY 1211 AGAGCTGTCCCAAGAGATGTCAGCGCTATGTGTGTACGAGAGGTCAACCAAGTCTGG 1270
 Db 1567 AGAGCTGTCCCAAGAGATGTCAGCGCTATGTGTGTACGAGAGGTCAACCAAGTCTGG 1626
 QY 1271 ATCAATGTTCAATGACATCTTCTATCCCTTCCCAATCAGAGGGAAGAGAGAGCTGTGC 1330
 Db 1627 ATCAATGTTCAATGACATCTTCTATCCCTTCCCAATCAGAGGGAAGAGAGAGCTGTGC 1686
 QY 1331 TTCTCTCGGCGCAATGATGAGAGACGCGCTTCTGCAATTTGTACAAAGTCACAGCGCTT 1390
 Db 1687 TTCTCTCGGCGCAATGATGAGAGACGCGCTTCTGCAATTTGTACAAAGTCACAGCGCTT 1746
 QY 1391 TTAATAATCCAGGCTACGATTTGAGTGAAGCCCTTACGCGCGGGAAGATGAATTTAAG 1450
 Db 1747 TTAATAATCCAGGCTACGATTTGAGTGAAGCCCTTACGCGCGGGAAGATGAATTTAAG 1806
 QY 1451 TGGCCCATTAAGGAAGAGATGTCTGTGACCAAGGAGTGAATGGAGAGTTTGGGAGGAC 1510
 Db 1807 TGGCCCATTAAGGAAGAGATGTCTGTGACCAAGGAGTGAATGGAGAGTTTGGGAGGAC 1866
 QY 1511 GGCTCCAAAGATGTGGGTCAATGAGAGACCAAGCTGTACTTCCAGGCGCACCAAGAGAC 1570
 Db 1867 GGCTCCA-----AGGCGACCAAGAGAC 1887
 QY 1571 ACGCCGCTGGAGACCACTCTACGTGTGACGTATGAGGCGGCGCGAGATCTGATCGC 1630
 Db 1888 ACGCCGCTGGAGACCACTCTACGTGTGACGTATGAGGCGGCGCGAGATCTGATCGC 1947

QY 1631 CTCACACGCGCGGCTTCCCATAGCTGTCTCATGAGGCGAGAACTTGACATGTTGCTC 1690
 Db 1948 CTCACACGCGCGGCTTCCCATAGCTGTCTCATGAGGCGAGAACTTGACATGTTGCTC 2007
 QY 1691 AGCCACTACAGCAGCGTGAAGCAGCGCGGCTGTGCTGACGTCTACAAAGCTGAGCGGCGC 1750
 Db 2008 AGCCACTACAGCAGCGTGAAGCAGCGCGGCTGTGCTGACGTCTACAAAGCTGAGCGGCGC 2067
 QY 1751 GACGACAGCCCCCTGCACAAGACGCCGCTTCTGGGCTAGCATGATGAGAGCGACGAC 1810
 Db 2068 GACGACAGCCCCCTGCACAAGACGCCGCTTCTGGGCTAGCATGATGAGAGCGACGAC 2127
 QY 1811 TGGCCCCGGATTAATGTTCTCTCAGAGATCTTCCATTTCCACAGCGGCTCGGATGTGCGG 1870
 Db 2128 TGGCCCCGGATTAATGTTCTCTCAGAGATCTTCCATTTCCACAGCGGCTCGGATGTGCGG 2187
 QY 1871 CTCCTACGATGATCTACAAAGCCCCACGCTTGCACGCGAGGAAGACACCCACGCTC 1930
 Db 2188 CTCCTACGATGATCTACAAAGCCCCACGCTTGCACGCGAGGAAGACACCCACGCTC 2247
 QY 1931 CTCCTTGTATGAGAGGCCCCCAGGTGACGCTGTGATTAATCTCTTCAAAAGGATCAAG 1990
 Db 2248 CTCCTTGTATGAGAGGCCCCCAGGTGACGCTGTGATTAATCTCTTCAAAAGGATCAAG 2307
 QY 1991 TACTTGGGCTCAACACACTGTGCTCCCTGTGGGTACGCGGTGTGTGATGACGCGAG 2050
 Db 2308 TACTTGGGCTCAACACACTGTGCTCCCTGTGGGTACGCGGTGTGTGATGACGCGAG 2367
 QY 2051 GGCTCTGTGACGAGAGGCTTGGTTCTGAAAGGGGCGCTGCAAAAACCAATGGGCGAGGT 2110
 Db 2368 GGCTCTGTGACGAGAGGCTTGGTTCTGAAAGGGGCGCTGCAAAAACCAATGGGCGAGGT 2427
 QY 2111 GAGATCAGAGACAGGTGAGAGGCTTGCAGTTCGTGTGCGGAGAAATGATGGCTTCATGAC 2170
 Db 2428 GAGATCAGAGACAGGTGAGAGGCTTGCAGTTCGTGTGCGGAGAAATGATGGCTTCATGAC 2487
 QY 2171 CTGAGCGAGGTTCGATTCATAGGCTGTGCTACGAGGCGCTCTCGCTCATGGGCGCTA 2230
 Db 2488 CTGAGCGAGGTTCGATTCATAGGCTGTGCTACGAGGCGCTCTCGCTCATGGGCGCTA 2547
 QY 2231 ATCCACAAGGCCCGAGTGTTCAGAGGTGCGCATGCGGGGTGCCCGGTCAACGCTGTGAGT 2290
 Db 2548 ATCCACAAGGCCCGAGTGTTCAGAGGTGCGCATGCGGGGTGCCCGGTCAACGCTGTGAGT 2607
 QY 2291 GCCTACGACACAGGTACACTGAGCGCTACATGAGAGTCCCTGAGAAACACAGCAGCGC 2350
 Db 2608 GCCTACGACACAGGTACACTGAGCGCTACATGAGAGTCCCTGAGAAACACAGCAGCGC 2667
 QY 2351 TATGAGGCGGCTCGGTGGCGCTGCAAGGTGAGAAAGCTGCGCAATGAGCGCAACGCGCTTG 2410
 Db 2668 TATGAGGCGGCTCGGTGGCGCTGCAAGGTGAGAAAGCTGCGCAATGAGCGCAACGCGCTTG 2727
 QY 2411 CTATCTCTCCAGCGCTTCTCGAGCAAAAGCTCACTTTTTCACACAAACTCTCTGCTC 2470
 Db 2728 CTATCTCTCCAGCGCTTCTCGAGCAAAAGCTCACTTTTTCACACAAACTCTCTGCTC 2787
 QY 2471 TCCCAACTGATTCGAGCAGGGAACCTTACAGCTCCAGATCTACCCCAAGAGAGACAC 2530
 Db 2788 TCCCAACTGATTCGAGCAGGGAACCTTACAGCTCCAGATCTACCCCAAGAGAGACAC 2847
 QY 2531 AGTATTGCGTCCCGAGTGGGCGAGCACTATGAAGTCACTGTGCACTTTCTACAG 2590
 Db 2848 AGTATTGCGTCCCGAGTGGGCGAGCACTATGAAGTCACTGTGCACTTTCTACAG 2907
 QY 2591 GAATACCTCTGAGC 2604
 Db 2908 GAATACCTCTGAGC 2921

RESULT 6
 ABR83338
 ID ABR83338 standard; cDNA; 4263 BP.

XX ABR83338;
AC 12-AUG-2002 (first entry)
XX CDNA encoding human DPRP-2 splice variant #6.
DE
XX
KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyslexia; reproductive disorder; inflammatory disorder;
KM metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX MO200231134-A2.
XX PN 18-APR-2002.
XX PD 12-OCT-2001; 2001WO-US31874.
XX PF 12-OCT-2001; 2000US-240117P.
XX PR 12-OCT-2000; 2000US-240117P.
XX PA (FERR) FERRING BV.
XX PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;
XX DR WPI: 2002-444178/47.
XX DR P-PSDB; ABR61607.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
XX -
XX Disclosure: Page 93-94; 113pp; English.
XX
PS The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
CC proteins (DPPR). The dipeptidyl peptidase IV (DPP-IV)-related
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting, or
CC psychologic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyslexias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABR83322-ABR83343 encode human DPRP proteins.
XX
SQ Sequence 4263 BP; 913 A; 1342 C; 1209 G; 799 T; 0 other;
Query Match 95.7%; Score 2504.4; DB 24; Length 4263;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 2554; Conservative 0; Mismatches 1; Indels 39; Gaps 1;

Db 547 AAGAGGATGAGTCTGGGCGCCACCTCCACCGCTCTACTACTGGGATGCAATATGGC 606
Oy 251 AGCCGAGAACTCCCTCTACTCTGATTTCCCAAGAGTCCGGAAGAAGCTCTG 310
Db 607 AGCCGAGAACTCCCTCTACTCTGATTTCCCAAGAGTCCGGAAGAAGCTCTG 666
Oy 311 CTGCTCTCTCTCTGGAAGAGATGCTGATCTTCAGAGCCAGCCACCAGTGGGTC 370
Db 667 CTGCTCTCTCTCTGGAAGAGATGCTGATCTTCAGAGCCAGCCACCAGTGGGTC 726
Oy 371 TACTCTCGGAGAGAGAGTCTGAGGAGCGGAAGACGCTTGGGCTCTTCCGATACC 430
Db 727 TACTCTCGGAGAGAGTCTGAGGAGCGGAAGACGCTTGGGCTCTTCCGATACC 786
Oy 431 TCTCTAGACTTTCACAGCAGAGAGTGGCCCTCTCTTCCAGAGCCAGCAACGCTTTC 490
Db 787 TCTCTAGACTTTCACAGCAGAGAGTGGCCCTCTCTTCCAGAGCCAGCAACGCTTTC 846
Oy 491 CACTGTCCGAGCGCGGCGCAAGAGCGCTTCATGTGTCCCTATGAAACGCTGAAATC 550
Db 847 CACTGTCCGAGCGCGGCGCAAGAGCGCTTCATGTGTCCCTATGAAACGCTGAAATC 906
Oy 551 AAGACCCAGTGTCTCAGAGGCGCCGATGACCCCAAAATTCGCTCCGACCTCTTC 610
Db 907 AAGACCCAGTGTCTCAGAGGCGCCGATGACCCCAAAATTCGCTCCGACCTCTTC 966
Oy 611 TCTCTCTCTATCAATTAACAGCAGCTGTGGTGGCCCAACATGAGAGAGGAGCGG 670
Db 967 TCTCTCTCTATCAATTAACAGCAGCTGTGGTGGCCCAACATGAGAGAGGAGCGG 1026
Oy 671 CGGCTGACCTTCTGCGACCAAGAGTTATCAATGTCTGATGACCCCAAGTCTCGGCT 730
Db 1027 CGGCTGACCTTCTGCGACCAAGAGTTATCAATGTCTGATGACCCCAAGTCTCGGCT 1086
Oy 731 GTGGCCACTTGTCTATACAGAGAGTTCGACCGCTTACTGTGGTGTGTGCCCC 790
Db 1087 GTGGCCACTTGTCTATACAGAGAGTTCGACCGCTTACTGTGGTGTGTGCCCC 1146
Oy 791 ACAGCCTCTGGAAGAGTTCAGAGGCGCTCAAGAGCTGCAATCTGTATGAGAATC 850
Db 1147 ACAGCCTCTGGAAGAGTTCAGAGGCGCTCAAGAGCTGCAATCTGTATGAGAATC 1206
Oy 851 GATGAGTCCGAGGTGAGGTCAATCACTGCTCTGCTGCTGAGAAAGAGAGAG 910
Db 1207 GATGAGTCCGAGGTGAGGTCAATCACTGCTCTGCTGCTGAGAAAGAGAGAG 1266
Oy 911 GACTGTATCTGATACCCAGAGCAGAGCAAGATTCACAAATGTGCTTGAATCTGCT 970
Db 1267 GACTGTATCTGATACCCAGAGCAGAGCAAGATTCACAAATGTGCTTGAATCTGCT 1326
Oy 971 GAGTTCAAGACTGAGAGCAGAGGCAAGATCTGTGACCCAGAGAGAGAGTGTGAG 1030
Db 1327 GAGTTCAAGACTGAGAGCAGAGGCAAGATCTGTGACCCAGAGAGAGAGTGTGAG 1386
Oy 1031 CCTTCACTGCTGTCTCCAGAGTGAATGATCCAGAGGCGCGGTGACCCGGGAT 1090
Db 1387 CCTTCACTGCTGTCTCCAGAGTGAATGATCCAGAGGCGCGGTGACCCGGGAT 1446
Oy 1091 GGCAATATACGCTTGGGCAATGTCTGAGACGCGCCGAGCAATGAGTCTCAGTCTGCTC 1150
Db 1447 GGCAATATACGCTTGGGCAATGTCTGAGACGCGCCGAGCAATGAGTCTCAGTCTGCTC 1506
Oy 1151 CTCCCGCGCGCTGTTCATCCAGAGCAGAGATGAGAGAGAGAGTCTAGCTCTGCTC 1210
Db 1507 CTCCCGCGCGCTGTTCATCCAGAGCAGAGATGAGAGAGAGAGTCTAGCTCTGCTC 1566
Oy 1211 AGAGTGTCCCGAGAGATGTCAGCGTATGTGTATAGAGAGAGTCAACAGCTTGG 1270
Db 1567 AGAGTGTCCCGAGAGATGTCAGCGTATGTGTATAGAGAGAGTCAACAGCTTGG 1626
Oy 1271 ATCAATGTTCATGACATCTTCTATCCCTTCCCAATCAGAGAGAGAGAGTCTGCTC 1330

```

Db 1627 ATCAATGTTATGACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGGACGAGCTGTC 1686
Qy 1331 TTTCTCCGCGCCATGAAATGCAAGACGGCTTTCGCCATTGTACAAAGTACCGCGCTT 1390
Db 1687 TTTCTCCGCGCCATGAAATGCAAGACGGCTTTCGCCATTGTACAAAGTACCGCGCTT 1746
Qy 1391 TTTAAATCCAGGGGCTACGATTTGGAGTGGAGCTTTCAGCCCGGGGGAAGATGAATTTAAG 1450
Db 1747 TTTAAATCCAGGGGCTACGATTTGGAGTGGAGCTTTCAGCCCGGGGGAAGATGAATTTAAG 1806
Qy 1451 TGCCCATTAAGAGAGATTTGCTCTGACACAGCGGTGAATGGAGGTTTGGCGAGGAC 1510
Db 1807 TGCCCATTAAGAGAGATTTGCTCTGACACAGCGGTGAATGGAGGTTTGGCGAGGAC 1866
Qy 1511 GGGTCCAAAGATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGGCAACAGGAC 1570
Db 1867 GGGTCCAA-----AGGGCAACCAAGGAC 1887
Qy 1571 AGCGCGGTGAGAGACACCTCTAGCTGGTCAAGTATGAGGGGGCGGAGATGCTAGCG 1630
Db 1888 AGCGCGGTGAGAGACACCTCTAGCTGGTCAAGTATGAGGGGGCGGAGATGCTAGCG 1947
Qy 1631 CTCACACAGCGCGGCTTCTCCATAGCTGTCTCCATGAGCCAGAACTTGCATGTTCGTC 1690
Db 1948 CTCACACAGCGCGGCTTCTCCATAGCTGTCTCCATGAGCCAGAACTTGCATGTTCGTC 2007
Qy 1691 AGCCACTACAGACGCTGAGACAGCGCGCTGCTGACGCTCTACAGCTGAGCGGCGCC 1750
Db 2008 AGCCACTACAGACGCTGAGACAGCGCGCTGCTGACGCTCTACAGCTGAGCGGCGCC 2067
Qy 1751 GAGGAGAGCGCGCTGAGAGAGCGCGCTTGGGCTAGCATGATGAGAGGACAGCGAC 1810
Db 2068 GAGGAGAGCGCGCTGAGAGAGCGCGCTTGGGCTAGCATGATGAGAGGACAGCGAC 2127
Qy 1811 TGCCCGCGGATATGTTCTCTCCAGAGATCTTCCATTTCCACACGCGTCCGATGTGCG 1870
Db 2128 TGCCCGCGGATATGTTCTCTCCAGAGATCTTCCATTTCCACACGCGTCCGATGTGCG 2187
Qy 1871 CTCTACGGCATGATTAACAAGCCCGCTTTCAGCCAGGAGAAAGACACCCACCGTC 1930
Db 2188 CTCTACGGCATGATTAACAAGCCCGCTTTCAGCCAGGAGAAAGACACCCACCGTC 2247
Qy 1931 CTCTTGTATATGAGGGCGCCAGGTGAGCTGATTAATCTCTTCAAGAGCATCAAG 1990
Db 2248 CTCTTGTATATGAGGGCGCCAGGTGAGCTGATTAATCTCTTCAAGAGCATCAAG 2307
Qy 1991 TACTTGGCGCTCAACACACTGGCTCCCTGGGCTACGGCTGATGATGAGCGCAGG 2050
Db 2308 TACTTGGCGCTCAACACACTGGCTCCCTGGGCTACGGCTGATGATGAGCGCAGG 2367
Qy 2051 GGGTCTCTGACGAGAGGGCTTCGTTGCAAGGGGCGCTGAAAACCAATGGGCCAGGTG 2110
Db 2368 GGGTCTCTGACGAGAGGGCTTCGTTGCAAGGGGCGCTGAAAACCAATGGGCCAGGTG 2427
Qy 2111 GAGATCGAGAGACAGGTGAGGGGCTGCACTGAGGGCGAAGATGAGCTTCAATCAGC 2170
Db 2428 GAGATCGAGAGACAGGTGAGGGGCTGCACTGAGGGCGAAGATGAGCTTCAATCAGC 2487
Qy 2171 CTGAGCCAGATGGCATCATGCTGCTGCTCCTACGGGGGCTTCTCTGCTCATGAGGGCTA 2230
Db 2488 CTGAGCCAGATGGCATCATGCTGCTGCTCCTACGGGGGCTTCTCTGCTCATGAGGGCTA 2547
Qy 2231 ATCCAAAGCCCGAGGTGTTCAAGGTGGCCATCGCGGGTGCCCGGCTACCGTCTGAGATG 2290
Db 2548 ATCCAAAGCCCGAGGTGTTCAAGGTGGCCATCGCGGGTGCCCGGCTACCGTCTGAGATG 2607
Qy 2291 GCGTACGACAGAGGATGACAGGCTGATGAGAGCTGCTCCGAGAAACCAAGACAGCGGCT 2350
Db 2608 GCGTACGACAGAGGATGACAGGCTGATGAGAGCTGCTCCGAGAAACCAAGACAGCGGCT 2667
Qy 2351 TATGAGGCGGGTTCCGTGGCGCTGACAGCTGAGAGAGCTGCCCAATGAGCCCAACGCTTG 2410
Db 2668 TATGAGGCGGGTTCCGTGGCGCTGACAGCTGAGAGAGCTGCCCAATGAGCCCAACGCTTG 2727

```

```

Qy 2411 CTFAATCCACAGGCTTCTGAGAGAAAGGTGACATTTTCCACACAACCTTCTGTC 2470
Db 2728 CTFAATCCACAGGCTTCTGAGAGAAAGGTGACATTTTCCACACAACCTTCTGTC 2787
Qy 2471 TCCCAACTGATCCAGAGAGGAAACCTTACAGCTCCAGATCTACCCCAAGAGACAC 2530
Db 2788 TCCCAACTGATCCAGAGAGGAAACCTTACAGCTCCAGATCTACCCCAAGAGAGACAC 2847
Qy 2531 AGTATTCGCTCCCGAGTGGGGGAGCAGTATGAAAGTACAGTTGCTGCACTTCTACAG 2590
Db 2848 AGTATTCGCTCCCGAGTGGGGGAGCAGTATGAAAGTACAGTTGCTGCACTTCTACAG 2907
Qy 2591 GAATACCTCTGAGC 2604
Db 2908 GAATACCTCTGAGC 2921

RESULT 7
AAD38957
ID AAD38957 standard; DNA; 2495 BP.
XX
AC AAD38957:
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 4 (DPP4)-like 2 DNA.
XX
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme; gene; ds.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 1..2493
FT CDS
FT
FT /*tag= a
FT /product= "Human DPP4-like 2 protein"
FT /transl_except= (pos: 703..705, aa:Gln)
FT /note= "CDS does not include start codon"
FT /partial
XX
PN WO200234900-A1.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-AU01388.
XX
PR 27-OCT-2000; 2000AU-0001078.
XX
PA (UNSY ) UNIV SYDNEY.
XX
PI Abbott CA, Gorell MD;
XX
DR WPI: 2002-454646/48.
XX
DR P-PSDB: AAE24171.
XX
PT New dipeptidyl peptidase (DPP) peptidases, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
PS Disclosure: Page 86-88; 91pp; English.
XX
CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP4-like 2 DNA.
XX
SO Sequence 2495 BP; 535 A; 783 C; 696 G; 481 T; 0 other:

```


Query Match 95.1%; Score 2488.6; DB 24; Length 2495;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2491; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 110 CTCGGAGCATTCACCGGACGCGCAATGACCTGGGCGCTCATGTCACAAAGGCGCC 169
 Db 1 CTCGGAGCATTCACCGGACGCGCAAGTACTCGGGCGCTCATGTCACAAAGGCGCC 60
 QY 170 CAGGACTTCCAGTTTGTGCAAGAGACGATGATGTGGGCCCACTCCACCGCTTAC 229
 Db 61 CAGGACTTCCAGTTTGTGCAAGAGACGATGATGTGGGCCCACTCCACCGCTTAC 120
 QY 230 TACCTGGGAATGCCAATATGCGACCGAGAACTCCCTCTCTACTCTGAGATTTCCAA 289
 Db 121 TACCTGGGAATGCCAATATGCGACCGGAGAACTCCCTCTCTACTCTGAGATTTCCAA 180
 QY 290 AAGTCCGGAAAGAGGTCTGCTCTCTCTCTCGAAGAGATGCTGATTCATTTCAG 349
 Db 181 AAGTCCGGAAAGAGGTCTGCTCTCTCTCTCGAAGAGATGCTGATTCATTTCAG 240
 QY 350 GCCACGCGCCACCATGAGGCTCTACTCTCGGAGAGAGAGTGTGAGAGCGGAAACGC 409
 Db 241 GCCACGCGCCACCATGAGGCTCTACTCTCGGAGAGAGAGTGTGAGAGCGGAAACGC 300
 QY 410 CTGGGGCTCTTGGGCAATCACCTCTACGACTTCCACAGCAGAGTGGCTCTTCTCTTC 469
 Db 301 CTGGGGCTCTTGGGCAATCACCTCTACGACTTCCACAGCAGAGTGGCTCTTCTCTTC 360
 QY 470 CAGGCGCAGCAAGCCCTTCCACTGTGCGAGCGCGCAAGAACGCTTCATGTGTCC 529
 Db 361 CAGGCGCAGCAAGCCCTTCCACTGTGCGAGCGCGCAAGAACGCTTCATGTGTCC 420
 QY 530 CCTATGAACCCGCTGGAATCAAGACCCAGTGTCTCAGGGCCCGGATGAGCCCAAAATC 589
 Db 421 CCTATGAACCCGCTGGAATCAAGACCCAGTGTCTCAGGGCCCGGATGAGCCCAAAATC 480
 QY 590 TGGCTGTGCGACCCCTGCTTCTTCTCTCTATCAATTAACAGCACTGTGGTGGCCAC 649
 Db 481 TGGCTGTGCGACCCCTGCTTCTTCTCTCTCAACAATTAACAGCACTGTGGTGGCCAC 540
 QY 650 ATGAGAGCAGGCGAGGAGGCGGCGTACCTTGTGCCACCAAGTTTATCCATGTCTG 709
 Db 541 ATGAGAGCAGGCGAGGAGGCGGCGTACCTTGTGCCACCAAGTTTATCCATGTCTG 600
 QY 710 GATGACCCCAAGTCTGCGGGTGTGGCCACTTGTCTATACAGGAAGTTCGACCGCTTC 769
 Db 601 GATGACCCCAAGTCTGCGGGTGTGGCCACTTGTCTATACAGGAAGTTCGACCGCTTC 660
 QY 770 ACTGGTACTGTGTGTCGCCACAGCCTCTGGGAGAGTTTCAGAGGCGCTCAAGAGCTG 829
 Db 661 ACTGGTACTGTGTGTCGCCACAGCCTCTGGGAGAGTTTCAGAGGCGCTCAAGAGCTG 720
 QY 830 CGAATCCTGTATGAGGAAGTGCATGATGCCAGGTGAGTCAATCACTCCCTCTCTCT 889
 Db 721 CGAATCCTGTATGAGGAAGTGCATGATGCCAGGTGAGTCAATCACTCCCTCTCTCT 780
 QY 890 GCGCTAGAAGAAAGAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 949
 Db 781 GCGCTAGAAGAAAGAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 840
 QY 950 AAGATTGCTTGAAGTGTGCTGATTCAGACTGACAGCCAGGCGCAAGATGCTCTGACC 1009
 Db 841 AAGATTGCTTGAAGTGTGCTGATTCAGACTGACAGCCAGGCGCAAGATGCTCTGACC 900
 QY 1010 CAGGAGAAAGAGTGTGTGCGACCCCTTCACTGCTTCCGGAAGTGGAGTCAATCGCC 1069
 Db 901 CAGGAGAAAGAGTGTGTGCGACCCCTTCACTGCTTCCGGAAGTGGAGTCAATCGCC 960
 QY 1070 AAGGCGCGGAGAGCGGAGTGGCAATAGCTTGGGCGCTTCTCTGAGCGGCGCCAG 1129
 Db 961 AAGGCGCGGAGAGCGGAGTGGCAATAGCTTGGGCGCTTCTCTGAGCGGCGCCAG 1020

QY 1130 CAGTGGCTCCAGCTGCTGCTCTCTCCCGCGGCGCTGTTCAATCCGAGACAGAGATGAG 1189
 Db 1021 CAGTGGCTCCAGCTGCTGCTCTCTCCCGCGGCGCTGTTCAATCCGAGACAGAGATGAG 1080
 QY 1190 GAGCAGCGCTACGCTCTGCGAGAGCTGCTCCAGGAATGCGGCTATGAGGTGATC 1249
 Db 1081 GAGCAGCGCTACGCTCTGCGAGAGCTGCTCCAGGAATGCGGCTATGAGGTGATC 1140
 QY 1250 GAGGAGTCCACCAAGCTGTGATCAATGTTCAATGATCTTCTATATCCCTTCCCAATCA 1309
 Db 1141 GAGGAGTCCACCAAGCTGTGATCAATGTTCAATGATCTTCTATATCCCTTCCCAATCA 1200
 QY 1310 GAGGAGAGAGAGCTCTGCTTCTCTCCGCGCAATGATGATGATGATGATGATGATG 1369
 Db 1201 GAGGAGAGAGAGCTCTGCTTCTCTCCGCGCAATGATGATGATGATGATGATGATG 1260
 QY 1370 TTGTACAAAGTCAACCGCGCTTAAATCCAGAGGCTACGATGAGTGAAGGCTTCAAGC 1429
 Db 1261 TTGTACAAAGTCAACCGCGCTTAAATCCAGAGGCTACGATGAGTGAAGGCTTCAAGC 1320
 QY 1430 CCCGGGAAGATGAATTTAAGTCCCATTAAGAAAGAGATGCTCTGACACAGCGTGA 1489
 Db 1321 CCCGGGAAGATGAATTTAAGTCCCATTAAGAAAGAGATGCTCTGACACAGCGTGA 1380
 QY 1490 TGGGAGGTTTGGCGAGGCAAGGCTCCAGAGTGTGGGTCAATGAGAGACCAAGCTGTG 1549
 Db 1381 TGGGAGGTTTGGCGAGGCAAGGCTCCAGAGTGTGGGTCAATGAGAGACCAAGCTGTG 1440
 QY 1550 TACTTCCAGGCGACCAAGGACAGCGCGCTGTGAGACACCTCTATGCTGTGCTATGAG 1609
 Db 1441 TACTTCCAGGCGACCAAGGACAGCGCGCTGTGAGACACCTCTATGCTGTGCTATGAG 1500
 QY 1610 GCGGCGGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1669
 Db 1501 GCGGCGGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1560
 QY 1670 CAGAACTTGACATGTTCTGTGACGCACTACAGAGCGGTGAGCAGCGCGCTGTGAC 1729
 Db 1561 CAGAACTTGACATGTTCTGTGACGCACTACAGAGCGGTGAGCAGCGCGCTGTGAC 1620
 QY 1730 GTCTACAAAGTGAAGCGCGCGCGCGAGACCGCTGACAAAGAGCGCGCTGTGAG 1789
 Db 1621 GTCTACAAAGTGAAGCGCGCGCGCGAGACCGCTGACAAAGAGCGCGCTGTGAG 1680
 QY 1790 AGCATGATGAGAGCAGCAGCTGCGCGCGCGGATGATGATGATGATGATGATGATG 1849
 Db 1681 AGCATGATGAGAGCAGCAGCTGCGCGCGCGGATGATGATGATGATGATGATGATG 1740
 QY 1850 CACAGCGCTGCGATGTGGGCTCTACGGCATGATCTACAGCCCGCAAGCTTTCAGCCA 1909
 Db 1741 CACAGCGCTGCGATGTGGGCTCTACGGCATGATCTACAGCCCGCAAGCTTTCAGCCA 1800
 QY 1910 GGGAGAGACCCCAAGCGCTCTTGTATGAGAGGCGCGCGAGTGGAGCTGTGAGAT 1969
 Db 1801 GGGAGAGACCCCAAGCGCTCTTGTATGAGAGGCGCGCGAGTGGAGCTGTGAGAT 1860
 QY 1970 AACTCTTCAAAAGCATCAAGTACTTGTGCGCTCAACACACTGCGCTCTCTGCGCTACGCG 2029
 Db 1861 AACTCTTCAAAAGCATCAAGTACTTGTGCGCTCAACACACTGCGCTCTCTGCGCTACGCG 1920
 QY 2030 GTGGTTGTATGAGCGGAGGCGCTCTGTCAAGAGGCGCTTGTGTAAGAGGCGCGCTG 2089
 Db 1921 GTGGTTGTATGAGCGGAGGCGCTCTGTCAAGAGGCGCTTGTGTAAGAGGCGCGCTG 1980
 QY 2090 AAAAACAATATGGGCGAGGTGAGATGAGAGACAGAGTGAAGGCGCTTGTGTAAGAGGCGCGCTG 2149
 Db 1981 AAAAACAATATGGGCGAGGTGAGATGAGAGACAGAGTGAAGGCGCTTGTGTAAGAGGCGCGCTG 2040
 QY 2150 GAGAAGTATGCTTCAATGACCTGAGCGAGCTGATGATGATGATGATGATGATGATGATG 2209
 Db 2041 GAGAAGTATGCTTCAATGACCTGAGCGAGCTGATGATGATGATGATGATGATGATGATG 2100
 QY 2210 TTCTCTGCTCATGAGGCGCTAATCCAAAGCCCAAGTGTTCAGAGGTGAGCATCGCGGCT 2269

Db 2101 TTCTCTCGCTCAATGGGGCTAATCCACAGAGCCAGGTTGTTCAAGGATGGCCATCGCGGGT 2160
QY 2270 GCCCGGCTACCGTCTGTGATGGCTTACGACACAGGGTACTGAGCCCTACATGAGCGTC 2329
Db 2161 GCGCCGGTCAACCGTCTGTGATGGCTTACGACACAGGGTACTGAGCCCTACATGAGCGTC 2220
QY 2230 CCTGAGAACACACAGCAGCGGCTATGAGCGGGGTTCCGTGGCCCTGACACGTGGAGAACCTG 2389
Db 2221 CCTGAGAACACACAGCAGCGGCTATGAGCGGGGTTCCGTGGCCCTGACACGTGGAGAACCTG 2280
QY 2290 CCCAATGAGCCCAACCGCTTGTATCTCCACAGCGCTTCCCTGACGAGAAACGTGCACTTT 2449
Db 2281 CCCAATGAGCCCAACCGCTTGTATCTCCACAGCGCTTCCCTGAGGAGAAACGTGCACTTT 2340
QY 2450 TTCCACACAAACTTCTCTGCTCTCCCAACTGATCCGACAGGAGAAACCTTACACCTCCAG 2509
Db 2241 TTCCACACAAACTTCTCTGCTCTCCCAACTGATCCGACAGGAGAAACCTTACACCTCCAG 2400
QY 2510 ATCTACCCCAACGAGACACAGTATCGCTGGCCGAGTCGGGCGAGCACTATGAAGTC 2569
Db 2401 ATCTACCCCAACGAGACACAGTATCGCTGGCCGAGTCGGGCGAGCACTATGAAGTC 2460
QY 2570 ACCTTGTGCACTTTCTACAGGAATACCTCTGAGC 2604
Db 2461 ACCTTGTGCACTTTCTACAGGAATACCTCTGAGC 2495

RESULT 8
ABK83337
ID ABK83337 standard; cDNA; 4076 BP.
XX
AC ABK83337;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPRP-2 splice variant #5.
XX
KW Human: serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyslexia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001MO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
XX
P-PSDB; ABG61606.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
XX
PS Disclosure; Page 90-91; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)

CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bullimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyslexias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX

SO Sequence 4076 BP; 879 A; 1276 C; 1143 G; 778 T; 0 other;

Query Match 87.7%; Score 2296.4; DB 24; Length 4076;
Best Local Similarity 94.4%; Pred. No. 0;
Matches 2450; Conservative 0; Mismatches 1; Indels 143; Gaps 1;

QY 11 ATGGCCACACCGGGACCCCAACGGCCGACGAGCGACGCGCCGACAGATGACCG 70
Db 367 ATGGCCACACCGGGACCCCAACGGCCGACGAGCGCGACGCGCCGACAGATGACCG 426
QY 71 GCGCCCGCTTCCAGGTGACAGAACTCGTGGAGCGGGCTCCGAGCATCATCACGCG 130
Db 427 GCGCCCGCTTCCAGGTGACAGAACTCGTGGAGCGGGCTCCGAGCATCATCACGCG 486
QY 131 AGCCGCAAGTACTGGGCTCATTTGTCAACAAGCGCCCAAGCACTTCCAGTTTGTGAC 190
Db 487 AGCCGCAAGTACTGGGCTCATTTGTCAACAAGCGCGCCCAAGCACTTCCAGTTTGTGAC 546
QY 191 AAGAGGATGATGTGGGGCCCACTCCACAGCGCTTACTACTCGGGGAATGCCATATGAGC 250
Db 547 AAGAGGATGATGTGGGGCCCACTCCACAGCGCTTACTACTCGGGGAATGCCATATGAGC 606
QY 251 AGCCGAGAACTCCCTCTCTACTGTAGATTCGAAAGAGTCCGAAAGAGGCTCTG 310
Db 607 AGCCGAGAACTCCCTCTCTACTGTAGATTCGAAAGAGTCCGAAAGAGGCTCTG 666
QY 311 CTGCTCTCTCTCTGAAAGACAGATCTGATCATTTCCAGGCGACGCCACCATGCGTC 370
Db 667 CTGCTCTCTCTCTGAAAGACAGATCTGATCATTTCCAGGCGACGCCACCATGCGTC 726
QY 371 TACTCTGGGAGGAGAGCTGCTGAGGAGGAGGAAAGCGGCGGGGGCTTCGGATACGC 430
Db 727 TACTCTGGGAGGAGAGCTGCTGAGGAGGAGGAAAGCGGCGGGGGCTTCGGATACGC 786
QY 431 TCTTACGACTTCCACAGCAGAGAGTGGCTTCTCTCTCCAGGCGACAGCCTCTTC 490
Db 787 TCTTACGACTTCCACAGCAGAGAGTGGCTTCTCTCTCCAGGCGACAGCCTCTTC 846
QY 491 CACTGTGCGCAGCGCGGACAGAGCGCTTCAATGTTGCCCTATGAAACCGCTGGAATC 550
Db 847 CACTGTGCGCAGCGCGGACAGAGCGCTTCAATGTTGCCCTATGAAACCGCTGGAATC 906
QY 551 AAGACCCAGTGTCCAGGGCCCGCGATGAGACCCCAAAATCTGCCGTGCAGCCTGCCTC 610
Db 907 AAGACCCAGTGTCCAGGGCCCGCGATGAGACCCCAAAATCTGCCGTGCAGCCTGCCTC 966
QY 611 TTCTCTTCAATCAATACAGCAGCTGTGGTGGCCAAATCGAGACAGCGGAGAGCGG 670
Db 967 TTCTCTTCAATCAATACAGCAGCTGTGGTGGCCAAATCGAGACAGCGGAGAGCGG 1026
QY 671 GCGCTGACCTTCTGCGCACCAAGTTTATCAATGTCTGATGACCCCACTGCGGGT 730
Db 1027 GCGCTGACCTTCTGCGCACCAAGTTTATCAATGTCTGATGACCCCACTGCGGGT 1086
QY 731 GTGGCAACCTTCTGATACAGAGAGTTCAGCCGCTTCACTGGGTACTGGTGGCC 790
Db 1087 GTGGCAACCTTCTGATACAGAGAGTTCAGCCGCTTCACTGGGTACTGGTGGCC 1146
QY 791 ACAGCTCTCTGGAAGTTTCAGAGCGCTCAAGACGCTGCATCTGTATGAGAAATC 850
Db 1147 ACAGCTCTCTGGAAGTTTCAGAGCGCTCAAGACGCTGCATCTGTATGAGAAATC 1206

PF 12-OCT-2001: 2001WO-US31874.
 XX
 PR 12-OCT-2000: 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI: 2002-444178/47.
 XX P-PSDB: AB661605.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 PS Disclosure: Page 87-88: 113pp: English.
 XX
 XX The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bullimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK8332-ABK8343 encode human DPRP proteins.
 XX
 SQ Sequence 4159 BP: 894 A; 1306 C; 1174 G; 785 T; 0 other:

Query Match 87.7%; Score 2296.4; DB 24; Length 4159;
 Best Local Similarity 94.4%; Pred. No. 0;
 Matches 2450; Conservative 0; Mismatches 1; Indels 143; Gaps 1;

QY 11 ATGGCCACACCGGACCCCAAGCGCGGACGAGCGACGCGCCACAGATGACCG 70
 DB 367 ATGGCCACACCGGACCCCAAGCGCGGACGAGCGCGCGCCACAGATGACCG 426
 QY 71 GCGCGCGGCTTCAGGAGCAAGACACGCGGAGCGGCGCGGAGCATGACAGCGC 130
 DB 427 GCGCGCGGCTTCAGGAGCAAGACACGCGGAGCGGCGCGGAGCATGACAGCGC 486
 QY 131 AGCGCAAGTACTGCGGCTCATTTGTCAACAAGCGCGCCACGATTCATTTGTGCGAG 190
 DB 487 AGCGCAAGTACTGCGGCTCATTTGTCAACAAGCGCGCCACGATTCATTTGTGCGAG 546
 QY 191 AAGACGATGAGTGTGGGCGCCACTCCACCGCTCTACTACTCTGGGAATCCATATGCG 250
 DB 547 AAGACGATGAGTGTGGGCGCCACTCCACCGCTCTACTACTCTGGGAATCCATATGCG 606
 QY 251 AGCGGAGGAACTCCCTCTCTACTCTGAGATTCACCAAGAGGTCGCGGAAGAGGCTGTG 310
 DB 607 AGCGGAGGAACTCCCTCTCTACTCTGAGATTCACCAAGAGGTCGCGGAAGAGGCTGTG 666
 QY 311 CTGCTCTCTCTCTGAGAGCAAGATGCTGATCATTTTCCAGGCGCGCCACCATATGGGGTTC 370
 DB 667 CTGCTCTCTCTCTGAGAGCAAGATGCTGATCATTTTCCAGGCGCGCCACCATATGGGGTTC 726
 QY 371 TACTCTGTGGGAGAGAGCTCTGAGGAGCGGAAAGCGCTGTGGGATCCAC 430
 DB 727 TACTCTGTGGGAGAGAGCTCTGAGGAGCGGAAAGCGCTGTGGGATCCAC 786
 QY 431 TCCTACGACTTCCACAGAGAGAGGCTCTTCTCTTCCAGGCGCGCAACAGCCTCTTC 490
 DB 787 TCCTACGACTTCCACAGAGAGAGGCTCTTCTCTTCCAGGCGCGCAACAGCCTCTTC 846
 QY 491 CACTGTGCGGAGCGGCGCAAGAGGCTTCTATGTGTCCCTATGAAACCGCTGGAATTC 550
 DB 1927 CACTGTGCGGAGCGGCGCAAGAGGCTTCTATGTGTCCCTATGAAACCGCTGGAATTC 1986

DB 847 CACTGTGCGGAGCGGCGCAAGAGGCTTCTATGTGTCCCTATGAAACCGCTGGAATTC 906
 QY 551 AAGACGATGAGTGTGGGCGCCCGGATGAGCCCAAAATTCGCTGTGCGAGCCGCTTC 610
 DB 907 AAGACGATGAGTGTGGGCGCCCGGATGAGCCCAAAATTCGCTGTGCGAGCCGCTTC 966
 QY 611 TTCTCTCTTCAATATACAGAGGCTGTGGTGGCCAAATCAGACGCGAGGAGCGG 670
 DB 967 TTCTCTCTTCAATATACAGAGGCTGTGGTGGCCAAATCAGAGGAGGAGGAGCGG 1026
 QY 671 CGGCTGACCTTGTGCGCACCAAGGTTATCCATATCTCTGATGACCCCAAGTCTGGGGT 730
 DB 1027 CGGCTGACCTTGTGCGCACCAAGGTTATCCATATCTCTGATGACCCCAAGTCTGGGGT 1086
 QY 731 GTGGCGACCTTGTGATACAGAGGAGTGTGACCGCTTCACTGGGTACTGGTGGTGGCC 790
 DB 1087 GTGGCGACCTTGTGATACAGAGGAGTGTGACCGCTTCACTGGGTACTGGTGGTGGCC 1146
 QY 791 ACACCTCTCTGGAAGGTTTCAGAGGCTCAAGACGCTGCAATCTGTATGAGGAATC 850
 DB 1147 ACACCTCTCTGGAAGGTTTCAGAGGCTCAAGACGCTGCAATCTGTATGAGGAATC 1206
 QY 851 GATGAGTCCGAGGTGAGGATTCACGCTCCCTCTCTGCGCTGAGAAAGAGAGACG 910
 DB 1207 GATGAGTCCGAGGTGAGGATTCACGCTCCCTCTCTGCGCTGAGAAAGAGAGACG 1266
 QY 911 GACTCGTATCGGTATCCCGACAGGACGAGCAAGATCCCAATTCGCTTAAACTGGCT 970
 DB 1267 GACTCGTATCGGTATCCCGACAGGACGAGCAAGATCCCAAGATTCGCTTAAACTGGCT 1326
 QY 971 GAGTTCAGAGTGTACAGCGCAGGCAAGATCTCTGACCCAGAGAGAGAGAGAGTGTGAG 1030
 DB 1327 GAGTTCAGAGTGTACAGCGCAGGCAAGATCTCTGACCCAGAGAGAGAGAGAGTGTGAG 1386
 QY 1031 CCTTCAGCTGCTGTTTCCGAGGTGAGTATCGGACGCGCGGTGAGCCCGGAT 1090
 DB 1387 CCTTCAGCTGCTGTTTCCGAGGTGAGTATCGGACGCGCGGTGAGCCCGGAT 1446
 QY 1091 GGCATAATACGCTGTGGCATTTCTGACCGCGCCAGCGAGTGGTCCACTGCTCTCC 1150
 DB 1447 GGCATAATACGCTGTGGCATTTCTGACCGCGCCAGCGAGTGGTCCACTGCTCTCC 1506
 QY 1151 CTCCCGCGGCTGTTTATCCGAGCACAGAGATGAGAGAGCGGCTAGCCTTCC 1210
 DB 1507 CTCCCGCGGCTGTTTATCCGAGCACAGAGATGAGAGAGCGGCTAGCCTTCC 1566
 QY 1211 AGAGCTGTCCCGAGGAATGTCCAGCGGTATGTGTGTAAGAGAGTCCACAGCTGTGG 1270
 DB 1567 AGAGCTGTCCCGAGGAATGTCCAGCGGTATGTGTGTAAGAGAGTCCACAGCTGTGG 1626
 QY 1271 ATCAATGTTCATGACATCTTATCCCTTCCCGCAATACAGAGGAGAGAGAGCTGTGC 1330
 DB 1627 ATCAATGTTCATGACATCTTATCCCTTCCCGCAATACAGAGGAGAGAGAGCTGTGC 1686
 QY 1331 TTCTCTCCGCGCAATGATGAGAGCGGCTTTCGCAATTTGTCAAAAGTACACGCGCTT 1390
 DB 1687 TTCTCTCCGCGCAATGATGAGAGCGGCTTTCGCAATTTGTCAAAAGTACACGCGCTT 1746
 QY 1391 TTAAATCCAGAGGCTACGATGTGAGTACGCTTTCACCGCGGGAAGATGATTTAAG 1450
 DB 1747 TTAAATCCAGAGGCTACGATGTGAGTACGCTTTCACCGCGGGAAGATGATTTAAG 1806
 QY 1451 TGCCCATTTAAGGAAGATGCTGTGACAGCGGTGTAATGGAGGTTTGGCGAGGAC 1510
 DB 1807 TGCCCATTTAAGGAAGATGCTGTGACAGCGGTGTAATGGAGGTTTGGCGAGGAC 1866
 QY 1511 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGCGCAACAGGAC 1570
 DB 1867 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGCGCAACAGGAC 1926
 QY 1571 AGCGCGGTGAGAGCACCTCTACGTGTCAAGTATGAGAGCGCGCGAGATCTGACG 1630
 DB 1927 AGCGCGGTGAGAGCACCTCTACGTGTCAAGTATGAGAGCGCGCGAGATCTGACG 1986

```
QY 1631 CTCACACGCGCGCTTCTCCATAGCTGCTCCATGAGCCAGAACTTGCATGTTGTC 1690
    |||||||
DB 1987 CTCACACGCGCGCTTCTCCATAGCTGCTCCATGAGCCAGAACTTGCATGTTGTC 2046
QY 1691 AGCCACTACAGCAGCGTGAGCAGCGCCCTGCTGTCAGCTTCAAGCTGAGCGGCC 1750
    |||||||
DB 2047 AGCCACTACAGCAGCGTGAGCAGCGCCCTGCTGTCAGCTTCAAGCTGAGCGGCC 2106
QY 1751 GAGGAGACCGCTGACAAAGCAGCGCCGCTTGGGCTGAGCTGAGTGGAGCGCAGC 1810
    |||||||
DB 2107 GAGGAGACCGCTGACAAAGCAGCGCCGCTTGGGCTGAGCTGAGTGGAGCGCAGC 2166
QY 1811 TGGCCCCGATTATGTTCTCTCCAGAGATCTTCCATTTCCACAGCGCTGAGTGGCG 1870
    |||||||
DB 2167 TGGCCCCGATTATGTTCTCTCCAGAGATCTTCCATTTCCACAGCGCTGAGTGGCG 2226
QY 1871 CTCTACGGCATGATCTACAAAGCCCGCTTGCAGCGCAGGAAAGACACCCCGCTC 1930
    |||||||
DB 2227 CTCTACGGCATGATCTACAAAGCCCGCTTGCAGCGCAGGAAAGACACCCCGCTC 2286
QY 1931 CTCTTTGATATGAGGCGCGCGCTGACAGCTGCTGAATACCTTCAAGGCATCAAG 1990
    |||||||
DB 2287 CTCTTTGATATGAGGCGCGCGCTGACAGCTGCTGAATACCTTCAAGGCATCAAG 2346
QY 1991 TACTTCGCGCTCAACACACTGCGCTCCCTGGCTACGCGCTGGTGTGATTGACGCGAG 2050
    |||||||
DB 2347 TACTTCGCGCTCAACACACTGCGCTCCCTGGCTACGCGCTGGTGTGATTGACGCGAG 2406
QY 2051 GCGCTCTGTACGAGGAGCGCTTCCGTTTGAAGGGGCCCTGAAAAACCAATGGCCAGGTG 2110
    |||||||
DB 2407 GCGCTCTGTACGAGGAGCGCTTCCGTTTGAAGGGGCCCTGAAAAACCAATGGCCAGGTG 2466
QY 2111 GAGATGAGGAGGAGGAGGAGGCGCTGCAAGTGTGCGGCGAGAAATGAGCTTCATCGAC 2170
    |||||||
DB 2467 GAGATGAGGAGGAGGAGGAGGCGCTGCAAGTGTGCGGCGAGAAATGAGCTTCATCGAC 2226
QY 2171 CTGAGCGGAGTTGCCATTCATTCATGCTGCTTACGAGGAGGCTTCTCTCGCATGAGGCTA 2230
    |||||||
DB 2527 CTGAGCGGAGTTGCCATTCATTCATGCTGCTTACGAGGAGGCTTCTCTCGCATGAGGCTA 2586
QY 2231 ATCCACAGCGCGAGGTGTTCAAGGTGGCCATGCGGGGTGCCCCGGTCACCGTCTGATG 2290
    |||||||
DB 2587 ATCCACAGCGCGAGGTGTTCAAG----- 2610
QY 2291 GCCTAGACACAGGATACACTGAGCGGTACATGAGAGTCCCTGAGAAACACGACGCGC 2350
    |||||||
DB 2611 ----- 2610
QY 2351 TATGAGCGGAGTTCCGTGGCCCTGACAGTGGAGAGCTGCCAATGAGCCCAACCGCTTG 2410
    |||||||
DB 2611 ----- 2610
QY 2411 CTATATCTCCACGCGCTTCTCTGACGAAAGCGTGCACATTTTTCACACAAACTTCTCTGTC 2470
    |||||||
DB 2624 CTATATCTCCACGCGCTTCTCTGACGAAAGCGTGCACATTTTTCACACAAACTTCTCTGTC 2683
QY 2471 TCCCACTGATCCGAGCAGGAGAAACCTTCCAGCTCAGATTCACCCCAAGGAGAGACAC 2530
    |||||||
DB 2684 TCCCACTGATCCGAGCAGGAGAAACCTTCCAGCTCAGATTCACCCCAAGGAGAGACAC 2743
QY 2531 AGTATTCGCTGCGCGAGTGGCGGAGCACTATGAAGTCACTGTTGCTGCACTTCTACAG 2590
    |||||||
DB 2744 AGTATTCGCTGCGCGAGTGGCGGAGCACTATGAAGTCACTGTTGCTGCACTTCTACAG 2803
QY 2591 GAATACCTCTGAGC 2604
    |||||||
DB 2804 GAATACCTCTGAGC 2817
    |||||||
```

RESULT 10
AA157896
ID AA157896 standard; cDNA: 2801 BP.

```
XX AC AA157896;
XX DT 22-OCT-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 99.
XX KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukaemia; ss.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000MO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR MPI: 2001-442253/47.
XX DR P-PSDB; AAM38740.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Claim 1, SEQ ID NO 99; 10078bp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC activating/inhibiting activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S. disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX XX
XX Sequence 2801 BP; 586 A; 889 C; 801 G; 525 T; 0 other:
XX
XX Query Match 87.4%; Score 2288.2; DB 22; Length 2801;
XX Best Local Similarity 94.1%; Pred. No. 0;
XX Matches 2485; Conservative 0; Mismatches 23; Indels 133; Gaps 5;
QY 11 ATGGCCACACGCGGAGCCCAACGAGCGGAGGAGGAGCGACCGCCAGATGAGCCG 70
    |||||||
DB 245 ATGGCCACACGCGGAGCCCAACGAGCGGAGGAGGAGCGACCGCCAGATGAGCCG 304
    |||||||
QY 71 GCGCGCGGCTTCAGGTGAGAGCACTGCTGGAGCGGCTCCGGAGCATCATCAGCGC 130
    |||||||
```

|||||
Db 305 GCCGCCGCTCCAGGTCGAGAAGCACTCGTGGAGCGGGCTCCGGACATCATCAGACGCG 364
QY 131 AGCGGAAGTACTCGGSCCTCATTTGCAACAAGCGCCGCCACGACTTCCAGTTGTGCAG 190
Db 365 AGCGGAAGTACTCGGSCCTCATTTGCAACAAGCGCCGCCACGACTTCCAGTTGTGCAG 424
QY 191 AAGACGATAGTCTGGGCCCCCACTCCACCGCTCTACTACCTGGGGAATGCCATATGGC 250
Db 425 AAGACGATAGTCTGGGCCCCCACTCCACCGCTCTACTACTGATTCGCCAAGAGGTCGGAAAGAGGCTCTG 484
QY 251 AGCGGAGAGAATCCTCCTCTCTACTCTGAGATTCCCAAGAAGTCCGAAAGAGGCTCTG 310
Db 485 AGCGGAGAGAATCCTCCTCTCTACTCTGAGATTCCCAAGAAGTCCGAAAGAGGCTCTG 544
QY 311 CTGCTCTCTCTCTGGAAGCAGATGCTGGATCATTTCCAGGCGACGCCCATGGGCTC 370
Db 545 CTGCTCTCTCTCTGGAAGCAGATGCTGGATCATTTCCAGGCGACGCCCATGGGCTC 604
QY 371 TACTCTCGGAGAGAGAGTGTGAGGAGCGGAACGCGTGGGGGCTTCGGCATACG 430
Db 605 TACTCTCGGAGAGAGAGTGTGAGGAGCGGAACGCGTGGGGGCTTCGGCATACG 664
QY 431 TCCTAGACTTCCACAGCGAGAGTGGCTCTCTCTTCCAGGCGACAGCAGCTCTTC 490
Db 665 TCCTAGACTTCCACAGCGAGAGTGGCTCTCTCTTCCAGGCGACAGCAGCTCTTC 724
QY 491 CACTGTGCGACGCGCGGCAAGAACGCTTCATG-----GTGTCC 529
Db 725 CACTGTGCGACGCGCGGCAAGAACGCTTCATGATGAGCCCTGTGCTGTCTC 784
QY 530 CCTATGAAACCGGTGGAATCAAGAACCCAGTGTCAAGGCGCCGGATGAGCCCAATATC 589
Db 785 CCTATGAAACCGGTGGAATCAAGAACCCAGTGTCAAGGCGCCGGATGAGCCCAATATC 844
QY 590 TGCCCTGCGACGCGCTGCTCTCTCTCTCTATCAATAAGCGAGCTGTGGGTGGCCAC 649
Db 845 TGCCCTGCGACGCGCTGCTCTCTCTCTCTATCAATAAGCGAGCTGTGGGTGGCCAC 904
QY 650 ATGAGACAGCGGAGAGAGCGCGGCTGACCTTCTGCCACCAAGGTTTATCCATGTCTG 709
Db 905 ATGAGACAGCGGAGAGAGCGCGGCTGACCTTCTGCCACCAAGGTTTATCCATGTCTG 964
QY 710 GATGACCCCAAGTCTGGGGGTGGGCCACCTGCTCATACAGAAAGTTTCAGCGGCTC 769
Db 965 GATGACCCCAAGTCTGGGGGTGGGCCACCTGCTCATACAGAAAGTTTCAGCGGCTC 1024
QY 770 ACTGTGACTGTGTGTGCCCCCAGAGCTCTGTGGAAAGTTTCAGAGGCGCTCAAGACGCTG 829
Db 1025 ACTGTGACTGTGTGTGCCCCCAGAGCTCTGTGGAAAGTTTCAGAGGCGCTCAAGACGCTG 1084
QY 830 CGAATCTGTATGAGAAAGTGTGATGAGTCCGAGGTGGAAGTCAATTCAGCTCTCTCT 889
Db 1085 CGAATCTGTATGAGAAAGTGTGATGAGTCCGAGGTGGAAGTCAATTCAGCTCTCTCT 1144
QY 890 GCGGTAGAAAGAGAGAGAGGAGTCTGTATCGGTACCCAGAGACAGGACAGCAAGATCCC 949
Db 1145 GCGGTAGAAAGAGAGAGAGGAGTCTGTATCGGTACCCAGAGACAGGACAGCAAGATCCC 1204
QY 950 AAGATTGCTTGAAGTGTGCTGAGTTCCAGATGACAGCCAGGGCAAGATCTGTGACC 1009
Db 1205 AAGATTGCTTGAAGTGTGCTGAGTTCCAGATGACAGCCAGGGCAAGATCTGTGACC 1264
QY 1010 CAGAGAGAAGAGAGTGTGTGACGCTCTTACGCTCGTGTTCGCCAAGGTGAGTACATGGC 1069
Db 1265 CAGAGAGAAGAGAGTGTGTGACGCTCTTACGCTCGTGTTCGCCAAGGTGAGTACATGGC 1324
QY 1070 AGGGCCGGGTGAGACCGGGATGGAATAGCCCTGGGCGCATGTCTCTGAGACGGGCCAG 1129
Db 1325 AGGGCCGGGTGAGACCGGGATGGAATAGCCCTGGGCGCATGTCTCTGAGACGGGCCAG 1384
QY 1130 CAGTGTCTCAGCTGTGTCTCTCCCGCGGCGCTGTTCATCCGAGACAGAGAAATGAG 1189
|||||

Db 1385 CAGTGTCTCAGCTGTGTCTCTCCCGCGGCGCTGTTCATCCGAGACAGAGAAATGAG 1444
QY 1190 GAGCA-GCGGCTAGGCTCTGCGAGAGCTGTCCCGAAGATGTCAGGCGATGTGGTGA 1248
Db 1445 GAGCAGCGGCTAGGCTCTGCGAGAGCTGTCCCGAAGATGTCAGGCGATGTGGTGA 1504
QY 1249 CGAGAGAGTACCAAGCTGTGATCATGTTCATGATCTCTTATCTCTTCCCATATC 1308
Db 1505 GAGAGAGTACCAAGCTGTGATCATGTTCATGATCTCTTATCTCTTCCCATATC 1564
QY 1309 AGAGGAGAGAGAGAGTGTGCTTCTCCGCCCAATGAATGAAGACGCGCTTCGCA 1368
Db 1565 AGAGGAGAGAGAGAGAGTGTGCTTCTCCGCCCAATGAATGAAGACGCGCTTCGCA 1624
QY 1369 TTTGTACAAAGTCAAGCGGCTTTTAAATCCAGAGGCTAGATTTGAGTGAAGCTTCAG 1428
Db 1625 TTTGTACAAAGTCAAGCGGCTTTTAAATCCAGAGGCTAGATTTGAGTGAAGCTTCAG 1684
QY 1429 CCGCGGGGAGATGAATTTAAGTCCCATTTAAGAAAGATTTGCTGTGACAGCGGTGA 1488
Db 1685 CCGCGGGGAGATGA-----1699
QY 1489 ATGGAGGTTTGTGCGAGGACGCGCTCCAAAGATCTGGGTCAATGAGAGACCAAGCTGT 1548
Db 1700 -----GCAGAGCCTGAGCAATGCTATCTGGGTCAATGAGAGACCAAGCTGT 1747
QY 1549 GTACTTCCAGGGACCAAGGACAGCGGCTGTGAGACACACCTCTAAGTGTGACTATGA 1608
Db 1748 GTACTTCCAGGGACCAAGGACAGCGGCTGTGAGACACACCTCTAAGTGTGACTATGA 1807
QY 1609 GCGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1668
Db 1808 GCGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1867
QY 1669 CCAGAACTTGACATGTTGCTGACGACATACAGAGCGGTGAGCAAGCGGCGCTGCTGCA 1728
Db 1868 CCAGAACTTGACATGTTGCTGACGACATACAGAGCGGTGAGCAAGCGGCGCTGCTGCA 1927
QY 1729 GGTGTACAGTGTAGCGGCG 1788
Db 1928 GGTGTACAGTGTAGCGGCG 1987
QY 1789 TAGCATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1848
Db 1988 TAGCATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2000
QY 1849 CCACAGCGGCTCGGATGTGGGCTCTACAGCGCATGATCTACAGGCCCAAGCGCTTGACGC 1908
Db 2021 CCACAGCGGCTCGGATGTGGGCTCTACAGCGCATGATCTACAGGCCCAAGCGCTTGACGC 2080
QY 1909 AGGGAAGAGACCCGCGCTCTTGTATGAGAGGCGCCCAAGGTGACAGTGTGGA 1968
Db 2081 AGGGAAGAGACCCGCGCTCTTGTATGAGAGGCGCCCAAGGTGACAGTGTGGA 2140
QY 1969 TAACTCTTCAAGAGCATCAAGTCTTGTGCGGCTCAACACACTGTGCGCTCCGCGCTACGC 2028
Db 2141 TAACTCTTCAAGAGCATCAAGTCTTGTGCGGCTCAACACACTGTGCGCTCCGCGCTACGC 2200
QY 2029 CGTGTGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2088
Db 2201 CGTGTGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2260
QY 2089 GAAAGCAATATGGGCGAGGTGAGATCGAGAGCAGAGTGAAGGCGCTGCAAGTGTGAGC 2148
Db 2261 GAAAGCAATATGGGCGAGGTGAGATCGAGAGCAGAGTGAAGGCGCTGCAAGTGTGAGC 2320
QY 2149 CGAAGATATGAGTCTATCGACTGAGCCGAGTGTGAGCATCATGAGTGTGAGTGTGAGTGT 2208
Db 2321 CGAAGATATGAGTCTATCGACTGAGCCGAGTGTGAGCATCATGAGTGTGAGTGTGAGTGT 2380
QY 2209 CTTCTCTCGCTCATGAGGAGTATATCAACAGCCCAAGTGTTCAGAGTGTGAGTGTGAGTGT 2268
Db 2381 CTTCTCTCGCTCATGAGGAGTATATCAACAGCCCAAGTGTTCAGAGTGTGAGTGTGAGTGT 2440
|||||


```

Db 1147 ACAGCCCTCGGGAGAGTTGAGAGGCGCTCAGACGCTGGAAATCCTGATAGAGAGTCT 1206
QY 851 GATGAGTCCCAAGGTGGAGGTCATTCACGTCCCTCTCTCCGCTGAGAGAAAGAAACAGC 910
Db 1207 GATGAGTCCCAAGGTGGAGGTCATTCACGTCCCTCTCTCCGCTGAGAGAAAGAAACAGC 1266
QY 911 GACTCGATGGTATACCCAGAGAGAGAGAGATCCCAAGATTGGCTTGAACCTGGCT 970
Db 1267 GACTCGATGGTATACCCAGAGAGAGAGAGATCCCAAGATTGGCTTGAACCTGGCT 1326
QY 971 GAGTCCAGACTGACAGCCAGGGGCAAGATGCTCGACCCAGAGAGAGAGAGAGAGAGAG 1030
Db 1327 GAGTCCAGACTGACAGCCAGGGGCAAGATGCTCGACCCAGAGAGAGAGAGAGAGAGAG 1386
QY 1031 CCTTCAGCTCGCTGTTCCCGAAGGTGAGTACATCCGACAGGGCGGGGTGAGACCCGGAT 1090
Db 1387 CCTTCAGCTCGCTGTTCCCGAAGGTGAGTACATCCGACAGGGCGGGGTGAGACCCGGAT 1446
QY 1091 GGCAGAAATACGCTGGGCGATGTTCTTGAGCGGGCCGAGAGTGGCTCCAGCTGCTCTC 1150
Db 1447 GGCAGAAATACGCTGGGCGATGTTCTTGAGCGGGCCGAGAGTGGCTCCAGCTGCTCTC 1506
QY 1151 CTCGCCCGGCGCTGTTCAATCCGAGCAGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 1210
Db 1507 CTCGCCCGGCGCTGTTCAATCCGAGCAGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 1566
QY 1211 AGAGCTGTCCCGAGAGATGTCAGCCGTATGTTGTTACGAGAGAGAGAGAGAGAGAGAG 1270
Db 1567 AGAGCTGTCCCGAGAGATGTCAGCCGTATGTTGTTACGAGAGAGAGAGAGAGAGAGAG 1626
QY 1271 ATCAATGTTATGATGATCTCTATCCCTTCGCCCAATCAGAGAGAGAGAGAGAGAGAG 1330
Db 1627 ATCAATGTTATGATGATCTCTATCCCTTCGCCCAATCAGAGAGAGAGAGAGAGAGAG 1686
QY 1331 TTTCTCCGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1390
Db 1687 TTTCTCCGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1746
QY 1391 TTTAAATCCCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1450
Db 1747 TTTAAATCCCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1806
QY 1451 TGCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1510
Db 1807 TGCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1866
QY 1511 GGTCCAGATCTGGGTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1570
Db 1867 GGTCCAGATCTGGGTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1887
QY 1571 ACGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1630
Db 1888 ACGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1947
QY 1631 CTCACAGAGCGCGCTCTCCATAGCTGCTCATGAGCCAGAACTTCGACATGTTGCTC 1690
Db 1948 CTCACAGAGCGCGCTCTCCATAGCTGCTCATGAGCCAGAACTTCGACATGTTGCTC 2007
QY 1691 AGCCACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1750
Db 2008 AGCCACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2067
QY 1751 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1810
Db 2068 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2127
QY 1811 TGCCCCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1870
Db 2128 TGCCCCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2187
QY 1871 CTCTAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1930

```

```

Db 2188 CTCTAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2247
QY 1931 CTCTAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1990
Db 2248 CTCTAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2307
QY 1991 TACTTGGGCTCAACACACTGAGCTCTCCCTGGGCTGAGAGAGAGAGAGAGAGAGAGAG 2050
Db 2308 TACTTGGGCTCAACACACTGAGCTCTCCCTGGGCTGAGAGAGAGAGAGAGAGAGAGAG 2367
QY 2051 GGTCTCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2110
Db 2368 GGTCTCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2427
QY 2111 GAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2170
Db 2428 GAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2487
QY 2171 CTGAGCGAGTGGCATATCATGAGCTGCTGCTACAGGGGAGCTTCTCTGCTATGAGGCTA 2230
Db 2488 CTGAGCGAGTGGCATATCATGAGCTGCTGCTACAGGGGAGCTTCTCTGCTATGAGGCTA 2547
QY 2231 ATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2290
Db 2548 ATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2571
QY 2291 GCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2350
Db 2572 ----- 2571
QY 2351 TATGAGGAGGAGTCCGTGCGCTGCTGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2410
Db 2572 -----GCCAACCGCTTG 2584
QY 2411 CTATGCTTCAAGAGGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2470
Db 2585 CTATGCTTCAAGAGGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2644
QY 2471 TCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2530
Db 2645 TCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2704
QY 2531 AGTATGCTGCTCCCGAGAGTGGGCGAGCAGTATGATGATGATGATGATGATGATGATG 2590
Db 2705 AGTATGCTGCTCCCGAGAGTGGGCGAGCAGTATGATGATGATGATGATGATGATGATG 2764
QY 2591 GAATACCTCTGAGC 2604
Db 2765 GAATACCTCTGAGC 2778

```

```

RESULT 12
ABR83340
ID ABR83340 standard; cDNA; 4120 BP.
XX
AC ABR83340;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DRRP-2 splice variant #8.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DRRP;
KW DRRP; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinetic; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200231134-A2.
XX

```


PD 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US31874.
XX
XX 12-OCT-2000; 2000US-240117P.
XX
XX (FERR) FERRING BV.
XX
XX Q1 S, Akinsanya KO, Riviere PJ, Junien J;
XX
XX MPI: 2002-444178/47.
XX P-PSDB; AB661609.
XX
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
XX the proteins, useful for treating e.g. fungal, bacterial, protozoan and
XX viral infections, cancers, allergies, neurological disorders, or pain
XX
XX -
XX
XX Disclosure: Page 100-101; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
XX proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
XX proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
XX and nucleic acids encoding them are useful for treating infections
XX such as fungal, bacterial, protozoan and viral infections, particularly
XX infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
XX pain, diabetes, precocious puberty, infertility, obesity, anorexia,
XX bulimia, Parkinson's disease, acute heart failure, hypotension,
XX hypertension, urinary retention, osteoporosis, angina pectoris,
XX stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
XX psychotic and neurological disorders (e.g. anxiety, dementia, or
XX schizophrenia), and dyskinesias. These may also be used in discovering
XX therapeutic agents for the treatment of reproductive, inflammatory and
XX metabolic disorders. ABR83322-ABR83343 encode human DPPR proteins.
XX
XX Sequence 4120 BP; 884 A; 1298 C; 1162 G; 776 T; 0 other:
XX
XX Query Match 84.4%; Score 2208.4; DB 24; Length 4120;
XX Best Local Similarity 92.9%; Pred. No. 0;
XX Matches 2411; Conservative 0; Mismatches 1; Indels 182; Gaps 2;
XX
QY 11 AAGGCGACACCGGGAGCCCAACGCGCGACGAGCGAGACGCCGCCACAGATGACCGG 70
DB 367 ATGGCCACACCGGGAGCCCAACGCGCGACGAGCGAGCGACGCCGCCACAGATGACCGG 426
QY 71 GCGCGCCGCTTCAGGTGAGAGCACTGCTGGAGCGGCTCCGAGCATCATCCAGCGC 130
DB 427 GCGCGCCGCTTCAGGTGAGAGCACTGCTGGAGCGGCTCCGAGCATCATCCAGCGC 486
QY 131 AGCCGCACTACTCGGGCTCATTTGCAACAAGCGGCCCGACGACTTCCAGTTTGTGCG 190
DB 487 AGCCGCACTACTCGGGCTCATTTGCAACAAGCGGCCCGACGACTTCCAGTTTGTGCG 546
QY 191 AAGAGGATGAGTCTGGGGCCCACTCCCAACGCGCTACTACTGGGAGTGCATTTGGC 250
DB 547 AAGAGGATGAGTCTGGGGCCCACTCCCAACGCGCTACTACTGGGAGTGCATTTGGC 606
QY 251 AGCCGAGAACTCCCTCCTTACTCTGAGATTCCCAAGAGTCCGGAAGAGGCTCTG 310
DB 607 AGCCGAGAACTCCCTCCTTACTCTGAGATTCCCAAGAGTCCGGAAGAGGCTCTG 666
QY 311 CTGCTCTCTCTCTGAGAGAGATGCTGATTCATTTCAGAGCCAGGCCACCATGGGGTC 370
DB 667 CTGCTCTCTCTCTGAGAGAGATGCTGATTCATTTCAGAGCCAGGCCACCATGGGGTC 726
QY 371 TACTCTCGGAGAGAGAGTCTGAGAGGAGCGGAAAGCGCTGGGCTCTTCGGCATACAC 430
DB 727 TACTCTCGGAGAGAGTCTGAGAGGAGCGGAAAGCGCTGGGCTCTTCGGCATACAC 786
QY 431 TCCATGACTTTCACAGCAGAGAGTGGCTTCTCTTCCAGAGCCAGCAAGCGCTTTC 490
DB 787 TCCATGACTTTCACAGCAGAGAGTGGCTTCTCTTCCAGAGCCAGCAAGCGCTTTC 846

QY 491 CACTGTGGGAGCGGGCGAAGAAAGGCTTTCAGTGTGCTCCCTATGAAACCGCTGAAATC 550
DB 847 CACTGTGGGAGCGGGCGAAGAAAGGCTTTCAGTGTGCTCCCTATGAAACCGCTGAAATC 906
QY 551 AAGACCACTGCTCAGAGGCGCCGGATGAGACCCCAAAATCTGCCCTCCAGCCTGCTTC 610
DB 907 AAGACCACTGCTCAGAGGCGCCGGATGAGACCCCAAAATCTGCCCTCCAGCCTGCTTC 966
QY 611 TTCTCTTCATCAATTAACAGCGACTGTGGTGGGCCAATCGAGACAGCGGAGAGCGG 670
DB 967 TTCTCTTCATCAATTAACAGCGACTGTGGTGGGCCAATCGAGACAGCGGAGAGCGG 1026
QY 671 CGGCTGACCTTCTGCGACCAAGGTTTATCCAAATGTCCTGATGACCCCAAGTCTGGGGT 730
DB 1027 CGGCTGACCTTCTGCGACCAAGGTTTATCCAAATGTCCTGATGACCCCAAGTCTGGGGT 1086
QY 731 GTGGGCACTTGTCTATACAGGAAGATTGACCGCTTCACTGGGTACTGTGTGCTCC 790
DB 1087 GTGGGCACTTGTCTATACAGGAAGATTGACCGCTTCACTGGGTACTGTGTGCTCC 1146
QY 791 ACAGCTCTGGGAAGTTTCAGAGGCTCAGAGCGCTGGAATCCTGTATGAGAAATC 850
DB 1147 ACAGCTCTGGGAAGTTTCAGAGGCTCAGAGCGCTGGAATCCTGTATGAGAAATC 1206
QY 851 GATGACTCCGAGGTGAGAGTCAATTCAGTCCCTCTCTGCTGCTAGAGAAAGAGAGAGC 910
DB 1207 GATGACTCCGAGGTGAGAGTCAATTCAGTCCCTCTCTGCTGCTAGAGAAAGAGAGAGC 1266
QY 911 GACTGTATTCGGTATCCCAAGAGACGACACAGAAATCCCAAGATTCTTGAATCTGGCT 970
DB 1267 GACTGTATTCGGTATCCCAAGAGACGACACAGAAATCCCAAGATTCTTGAATCTGGCT 1326
QY 971 GAGTTCCAGACTGACAGCGGAGGCAAGATCTGCTGACCCAGAGAGAGAGAGTGGTGCAG 1030
DB 1327 GAGTTCCAGACTGACAGCGGAGGCAAGATCTGCTGACCCAGAGAGAGAGAGTGGTGCAG 1386
QY 1031 CCGTTTCAGCTGCTGTTCCCGAAGGTGAGTACATGCCAGGCGCGGTGAGACCGCGGAT 1090
DB 1387 CCGTTTCAGCTGCTGTTCCCGAAGGTGAGTACATGCCAGGCGCGGTGAGACCGCGGAT 1446
QY 1091 GCGAATATGCGCTGGGCGCAATGTTCTGTGACCGGCGCCAGACAGTGGTCCAGCTCGTCTC 1150
DB 1447 GCGAATATGCGCTGGGCGCAATGTTCTGTGACCGGCGCCAGACAGTGGTCCAGCTCGTCTC 1506
QY 1151 CTCGCCCGGCGCTGTTCTCCGAGAGGTGAGTACATGCCAGGCGCGGTGAGACCGCGGAT 1210
DB 1507 CTCGCCCGGCGCTGTTCTCCGAGAGGTGAGTACATGCCAGGCGCGGTGAGACCGCGGAT 1566
QY 1211 AGAGCTTCCCAAGAGATTCACAGCCGTATGTGTGTACAGAGAGGTTCACCAAGCTGTG 1270
DB 1567 AGAGCTTCCCAAGAGATTCACAGCCGTATGTGTGTACAGAGAGGTTCACCAAGCTGTG 1626
QY 1271 ATCAATGTTCAATGACATCTTATCCCTTCCCAATGAGAGGAGAGAGAGAGTCTGTC 1330
DB 1627 ATCAATGTTCAATGACATCTTATCCCTTCCCAATGAGAGGAGAGAGAGAGTCTGTC 1686
QY 1331 TTTTCTCCGCGCAATGAATGACAAAGCGGCTTCTGCAATTTGTACAAATTCACCGCGTT 1390
DB 1687 TTTTCTCCGCGCAATGAATGACAAAGCGGCTTCTGCAATTTGTACAAATTCACCGCGTT 1746
QY 1391 TTTTAAATCCAGAGGCTACGATTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 1450
DB 1747 TTTTAAATCCAGAGGCTACGATTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 1806
QY 1451 TGCCTCATTAAGAGAGATGCTCTGACACAGCGGTGAATGGAGAGTGGAGTGGAGAGC 1510
DB 1807 TGCCTCATTAAGAGAGATGCTCTGACACAGCGGTGAATGGAGAGTGGAGTGGAGAGC 1866
QY 1511 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGAGGACCAAGAGAC 1570
DB 1867 GGCCTCA-----AGGACCAAGAGAC 1887
QY 1571 AGCGCGTGGAGACCACTCTACGTGTGATGAGAGCGCGCGGAGATCGTATCGC 1630

Db	1888	ACGGCGGTGGAGACACACCTCTACAGTGGTGACACTATGAGGGCGCGGGAGATGGTAAGC	1947
Qy	1631	CTCACACAGCGCGGGCTTCTCCCATAGCTGCTCCATGAGCCAGAACCTTGACATGTTGTC	1690
Db	1948	CTCACACAGCGCGGGCTTCTCCCATAGCTGCTCCATGAGCCAGAACCTTGACATGTTGTC	2007
Qy	1691	AGCCACTACAGCAGCGGTGAGCAGCGCGCGCTCGGTGACAGTCTACAACCTGAGGGCCCC	1750
Db	2008	AGCCACTACAGCAGCGGTGAGCAGCGCGCGCTCGGTGACAGTCTACAACCTGAGGGCCCC	2067
Qy	1751	GAGCAGCAGCCCCCTGACAAAGCAGCCCCGCTTCTGGAGTAGCATGATGAGGAGCAGCCAGC	1810
Db	2068	GAGCAGCAGCCCCCTGACAAAGCAGCCCCGCTTCTGGAGTAGCATGATGAGGAGCAGCCAGC	2127
Qy	1811	TGCCCCCGGATTATGTTCCTCCAGAGATCTTTCATTTCCACACGCGCTCGGATGTGCG	1870
Db	2128	TGCCCCCGGATTATGTTCCTCCAGAGATCTTTCATTTCCACACGCGCTCGGATGTGCG	2187
Qy	1871	CTCTACGGCATGATCTACAAAGCCCCCAGCGCTTCTGACCCAGGAGAAAGACACCCACGCTC	1930
Db	2188	CTCTACGGCATGATCTACAAAGCCCCCAGCGCTTCTGACCCAGGAGAAAGACACCCACGCTC	2247
Qy	1931	CTCTTTGATATGAGAGCCCCCAGGAGCAGCTGAGTAATCACTCTTCAAGGCACTCAAG	1990
Db	2248	CTCTTTGATATGAGAGCCCCCAGGAGCAGCTGAGTAATCACTCTTCAAGGCACTCAAG	2307
Qy	1991	TACTTGGCGCTCAACACACTGCGGCTCCCTGAGGCTACGCGCTGGTGTGATTTAGCAGGCAAG	2050
Db	2308	TACTTGGCGCTCAACACACTGCGGCTCCCTGAGGCTACGCGCTGGTGTGATTTAGCAGGCAAG	2367
Qy	2051	GGCTCTCTGTACAGGAGAGGGCTTCGGTTTCGAAAGGGGCGCTGAAACCAAAATGGGCCAGGTG	2110
Db	2368	GGCTCTCTGTACAGGAGAGGGCTTCGGTTTCGAAAGGGGCGCTGAAACCAAAATGGGCCAGGTG	2427
Qy	2111	GAGATGAGAGAACAGGTGAGGGGCGTCGAGTTCTGAGGCGGAGAACTATAGGCTTCACTGAC	2170
Db	2428	GAGATGAGAGAACAGGTGAGGGGCGTCGAGTTCTGAGGCGGAGAACTATAGGCTTCACTGAC	2487
Qy	2171	CTGAGCGGATTTGCCATTCATGCGCTGGTCTTACAGGGGGCTTCTCTGCTCATGGAGGCTA	2230
Db	2488	CTGAGCGGATTTGCCATTCATGCGCTGGTCTTACAGGGGGCTTCTCTGCTCATGGAGGCTA	2547
Qy	2231	ATCCACAAGCCCCAGGTGTTCAAAGGTGGCCATTCGCGGTTGCCCGGGTCAACGCTGTGATG	2290
Db	2548	ATCCACAAGCCCCAGGTGTTCAAAGGTGGCCATTCGCGGTTGCCCGGGTCAACGCTGTGATG	2571
Qy	2291	GCTTCAGACACAGGGGTACACTGAGCGCTACATGAGAGTCCCTGAGAACACACAGCAGCGC	2350
Db	2572	GCTTCAGACACAGGGGTACACTGAGCGCTACATGAGAGTCCCTGAGAACACACAGCAGCGC	2571
Qy	2351	TATGAGCGGGATTCCGTTGGGCTTCGACAGTGGAGAAAGCTGCCAATGAGGCCAACCGCTTG	2410
Db	2572	TATGAGCGGGATTCCGTTGGGCTTCGACAGTGGAGAAAGCTGCCAATGAGGCCAACCGCTTG	2584
Qy	2411	CTTATCTCCACAGGCTTCTGAGCAGAAAGCTGACATTTTTCACAGCAAACTTCTCGTC	2470
Db	2585	CTTATCTCCACAGGCTTCTGAGCAGAAAGCTGACATTTTTCACAGCAAACTTCTCGTC	2644
Qy	2471	TCCCAACTGATCCGAGCAGGAGAAACCTTACAGAGCTCAGATTTACCCCAAGAGAGACAC	2530
Db	2645	TCCCAACTGATCCGAGCAGGAGAAACCTTACAGAGCTCAGATTTACCCCAAGAGAGACAC	2704
Qy	2531	AGTATTTGGTCCCCGAGTGGGGCGAGAGCACTATGAAGTACAGTCTGTGACTTTCTTACAG	2590
Db	2705	AGTATTTGGTCCCCGAGTGGGGCGAGAGCACTATGAAGTACAGTCTGTGACTTTCTTACAG	2764
Qy	2591	GAATACCTGTGAGC 2604	
Db	2765	GAATACCTGTGAGC 2778	

ID	AAI57880
XX	AAI57880 standard; cDNA: 3262 BP.
AC	
XX	AAI57880;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 83.
XX	
KW	Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
PD	
XX	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0532317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
PA	(HYSE-) HYSEQ INC.
XX	
PL	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
P1	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
DR	WPI: 2001-442253/47.
XX	P-PsDB: AAM38724..
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PS	such as central nervous system injuries -
XX	
PS	Claim 1; SEQ ID NO 83; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57796-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SO	Sequence 3262 BP; 687 A; 1019 C; 931 G; 625 T; 0 other;
XX	
Query Match	79.9%; Score 2091.4; DB 22; Length 3262;
Best Local Similarity	94.4%; Pred. No. 0;
Matches 2256; Conservative	0; Mismatches 21; Indels 112; Gaps 4;
OY	243 CATATGCAAGCCGAAGAAGTCTTCCTACTGTGAGTAFTCCCAAGAAAGTGCGGAAG 302 33 CATATGGCAGCCGAAGAAGTCTTCCTACTGTGAGTAFTCCCAAGAAAGTGCGGAAG 91

QY 303 AGGCTGCTGCTCTGCTCTGCTGGAAGCAGATGCTGATCATTTTCAGAGCCACGCCACC 362
|||||
Db 92 AGGCTGCTGCTGCTCTGCTCTGGAAGCAGATGCTGATCATTTTCAGAGCCACGCCACC 151
QY 363 ATGGGGTCTACTCTCGGGAGAGAGAGACTGCTGAGGAGCGGAAACGCTGGGGTCTTG 422
|||||
Db 152 ATGGGGTCTACTCTCGGGAGAGAGAGACTGCTGAGGAGCGGAAACGCTGGGGTCTTG 211
QY 423 GCATCACCTCTCTGACACTTCCACAGGAGAGAGTGGCTCTCTCTCTCCAGAGCCAGCAAC 482
|||||
Db 212 GCATCACCTCTCTGACACTTCCACAGGAGAGAGTGGCTCTCTCTCTCCAGAGCCAGCAAC 271
QY 483 GCCTCTTTCACCTGTCGACAGCGGCGCAAGAACGGCTTCATGCTGTCCTCCCTATGAAACCG 542
|||||
Db 272 GCCTCTTTCGCTGCTCCGACAGCGGCGCAAGAACGGCTTCATGCTGTCCTCCCTATGAAACCG 311
QY 543 TGGAAATCAAGACCCAGTCTGAGGGCCCCGGATGAGACCCCAAAATCTGCCCTCCGAC 602
|||||
Db 332 TGGAAATCAAGACCCAGTCTGAGGGCCCCGGATGAGACCCCAAAATCTGCCCTCCGAC 391
QY 603 CTGCTCTTCTCTCTCATCATCAATTAAGAGGACCTGTGGGTGGCCAAATCGAGACAGCG 662
|||||
Db 392 CTGCTCTTCTCTCTCATCATCAATTAAGAGGACCTGTGGGTGGCCAAATCGAGACAGCG 451
QY 663 AGGAGCGGCGGCTGACCTTCTGCGACCAAGTTTATCCAAATGTCGTGATGACCCCAAGT 722
|||||
Db 452 AGGAGCGGCGGCTGACCTTCTGCGACCAAGTTTATCCAAATGTCGTGATGACCCCAAGT 511
QY 723 CTGCGGGTGGGCGACCTTCTGTCATACAGAAAGTTTCAACCGCTTCACTGGGTACTGG 782
|||||
Db 512 CTGCGGGTGGGCGACCTTCTGTCATACAGAAAGTTTCAACCGCTTCACTGGGTACTGG 571
QY 783 GGTGCCCAAGCCTCTCTGGAAAGTTTCAGAGGGCTCAAGAGCTGCGAATCTCTATG 842
|||||
Db 572 GGTGCCCAAGCCTCTCTGGAAAGTTTCAGAGGGCTCAAGAGCTGCGAATCTCTATG 631
QY 843 AGGAAGTCATGATGATCCGAGGTGAGGTCAATTCAGTCCCTCTCTCTGCGCTAGAAAGAA 902
|||||
Db 632 AGGAAGTCATGATGATCCGAGGTGAGGTCAATTCAGTCCCTCTCTCTGCGCTAGAAAGAA 691
QY 903 GGAAGAGGAGCTGATGCTGCTACCCAGAGACAGAGGACGAAGATCCCAAGATTCGCTTGA 962
|||||
Db 692 GGAAGAGGAGCTGATGCTGCTACCCAGAGAGGACGAAGATTCCAAGATTCGCTTGA 751
QY 963 AACTGCTGATGTTCCAGATGACAGCCAGAGGCAAGATCTCTCGACCCAGAGAAAGAGC 1022
|||||
Db 752 AACTGCTGATGTTCCAGATGACAGCCAGAGGCAAGATCTCTCGACCCAGAGAAAGAGC 811
QY 1023 TGGTGCAGCCCTTCAGCTGCTGTTCCCGAAGGTGAGTACATCGCCAGAGGCCGGTGA 1082
|||||
Db 812 TGGTGCAGCCCTTCAGCTGCTGTTCCCGAAGGTGAGTACATCGCCAGAGGCCGGTGA 871
QY 1083 CCCGGATGGCAAAATACGCTGGGCGCATGTCCTGAGACGGGCCCCAGAGTGGCTCAGC 1142
|||||
Db 872 CCCGGATGGCAAAATACGCTGGGCGCATGTCCTGAGACGGGCCCCAGAGTGGCTCAGC 931
QY 1143 TGGTCTCTCTCCCGCGGCTGTTTCATCCCGAGCAGAGAGATGAGAGACGCGCTAG 1202
|||||
Db 932 TGGTCTCTCTCCCGCGGCTGTTTCATCCCGAGCAGAGAGATGAGAGACGCGCTAG 991
QY 1203 CCTGTGCGAGAGCTGCCAGAGATGTCAGACCGTATGTTGTTGTAAGAGAGTCA 1262
|||||
Db 992 CCTGTGCGAGAGCTGCCAGAGATGTCAGACCGTATGTTGTTGTAAGAGAGTCA 1051
QY 1263 AGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1322
|||||
Db 1052 AGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1111
QY 1323 AGCTGTGCTTTCGCGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1382
|||||
Db 1112 AGCTGTGCTTTCGCGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1171

QY 1383 CGCGCGTTTTAAATCCAGGGGCTAGCATTTGAGTGAAGCCCTTCACGCCCGGGGAAGATG 1442
|||||
Db 1172 CGCGCGTTTTAAATCCAGGGGCTAGCATTTGAGTGAAGCCCTTCACGCCCGGGGAAGATG 1231
QY 1443 AATTTAAGTCCCATTTAAGGAAGATTTGCTGTACACAGCGGTGAATGGAGGTTTGG 1502
|
Db 1232 A----- -cc 1234
QY 1503 CGAGCAGCGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTTCAGAGCA 1562
|||
Db 1235 AGAGCCTGACGAATGCTATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTTCAGAGCA 1294
QY 1563 CCAAGGACAGCGCGCTGAGACACACCTCTACGTGTACGTATGAGGCGCGCGGAGA 1622
|||||
Db 1295 CCAAGGACAGCGCGCTGAGACACACCTCTACGTGTACGTATGAGGCGCGCGGAGA 1354
QY 1623 TGGTACGCTCAACAGCGCGGCTTCCCATGAGTGTCTCCATGAGCCGCAACTTGAGCA 1682
|||||
Db 1355 TGGTACGCTCAACAGCGCGGCTTCCCATGAGTGTCTCCATGAGCCGCAACTTGAGCA 1414
QY 1683 TGTTCGTACGCCACTACAGCAGCGGTGAGACAGCGCGCTGCGTGCACGTCTACAACTGA 1742
|||||
Db 1415 TGTTCGTACGCCACTACAGCAGCGGTGAGACAGCGCGCTGCGTGCACGTCTACAACTGA 1474
QY 1743 GCGGCCCCGACGACGACCCCTGACACAGACGCCGCTTCTGGGCTAGCATGATGAGG 1802
|||||
Db 1475 GCGGCCCCGACGACGACCCCTGACACAGACGCCGCTTCTGGGCTAGCATGATGAGG 1534
QY 1803 CAGCCAGCTGCCCCCGGAAATATGTTCTGTCAGAGATGTTCCATTTCCACAGCGGCTGG 1862
|||||
Db 1535 CAGCCA----- -AGATCTTCCATTTTCCACAGCGGCTGG 1567
QY 1863 ATGTGCGGCTCTACGCGATGATCTACAAAGCCCGACCTTGGACCGCAGGGAAGAAACAC 1922
|||||
Db 1568 ATGTGCGGCTCTACGCGATGATCTACAAAGCCCGACCTTGGACCGCAGGGAAGAAACAC 1627
QY 1923 CCAACGCTCTCTTGTATATGAGAGGCCCGCAGGTGACGTGTGTAATTAACCTCTCAAG 1982
|||||
Db 1628 CCAACGCTCTCTTGTATATGAGAGGCCCGCAGGTGACGTGTGTAATTAACCTCTCAAG 1687
QY 1983 GCATCAAGTATGAGGCGCTCAACACACATGCGCTCCCTGGGCTACGCGGTGATGATG 2042
|||||
Db 1688 GCATCAAGTATGAGGCGCTCAACACACATGCGCTCCCTGGGCTACGCGGTGATGATG 1747
QY 2043 ACAGGAGGGGCTCTGTCAGCGAGGGCTTCGGTTCGAAGGGGCCCTGAAGAAACCAAAATG 2102
|||||
Db 1748 ACAGGAGGGGCTCTGTCAGCGAGGGCTTCGGTTCGAAGGGGCCCTGAAGAAACCAAAATG 1807
QY 2103 GCCAGGTGAGATCGAGACACAGGTGAGGGCTTGCAGTTCTGTGGCGGAAGTATGCT 2162
|||||
Db 1808 GCCAGGTGAGATCGAGACACAGGTGAGGGCTTGCAGTTCTGTGGCGGAAGTATGCT 1867
QY 2163 TCATGACCTGAGACCGAGTGGCATCATGAGTGGTCTACAGGGGGCTTCCTGCTCTCA 2222
|||||
Db 1868 TCATGACCTGAGACCGAGTGGCATCATGAGTGGTCTACAGGGGGCTTCCTGCTCTCA 1927
QY 2223 TGGGGCTAAATCCAAAGCCCAAGGTGTTCAAGGTGGCCATTCGCGGGTCCCGGCTACCG 2282
|||||
Db 1928 TGGGGCTAAATCCAAAGCCCAAGGTGTTCAAGGTGGCCATTCGCGGGTCCCGGCTACCG 1987
QY 2283 TGTGATGGCTTACGACACAGGGTACACTGAGCGCTACATGGAAGCTCTGTGAGAACAC 2342
|||||
Db 1988 TGTGATGGCTTACGACACAGGGTACACTGAGCGCTACATGGAAGCTCTGTGAGAACAC 2047
QY 2343 AGCAGGCTATGAGGCGGGTTCGGTCCGTGAGTGGAGAAAGTGGCCATGAGGCCCA 2402
|||||
Db 2048 AGCAGGCTATGAGGCGGGTTCGGTCCGTGAGTGGAGAAAGTGGCCATGAGGCCCA 2107
QY 2403 ACCGTTGCTTATCTCTCAGCGCTTCTGAGAGAAACCTGACATTTTTCACACAAACT 2462
|||||
Db 2108 ACCGTTGCTTATCTCTCAGCGGCTTCTGAGAGAAACCTGACATTTTTCACACAAACT 2167
QY 2463 TCTTGCTCTCCCAACTGATCCGACAGGGAACCTTACAGCTC----- 2506

|||||
Db 2168 TCCTGCTCCCAACTGATCCGAGCAGGAACTTACCAAGTCCAGGTGGCCCTGCTC 2227
QY 2507 -----CAGATCTACCCCAAGAGAGACAGATATCCCTCCCGAATCGGGC 2555
Db 2228 CTGCTCCCGCGAGCTACCCCAAGAGAGACAGATATCCCTCCCGAATCGGGC 2287
QY 2556 AGCACTATGAGTACAGTCTGCTCAGCTTCTACAGAAATACCTCTGAGC 2604
Db 2288 AGCACTATGAGTACAGTCTGCTCAGCTTCTACAGAAATACCTCTGAGC 2336

RESULT 14
AAD38955
ID AAD38955 standard: cDNA; 3287 BP.
AC AAD38955;
XX
XX
XX 23-SEP-2002 (first entry)
XX
XX
DE Alternative version of murine dipeptidyl peptidase 9 (DPP9) cDNA.
XX
XX Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
KW human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
KW type II diabetes; antidiabetic; antinflammatory; immunosuppressive;
KW antiviral; enzyme; gene; ss.
XX
XX Mus sp.
OS
XX
XX
XX Key Location/Qualifiers
FH 1..2610
FT CDS /tag= a
FT /product= "Murine DPP9 protein"
FT /note= "CDS does not include start codon"
FT /partial
XX
XX WO200234900-A1.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-AU01388.
XX
XX 27-OCT-2000; 2000AU-0001078.
XX
XX (UNSY) UNITV SYDNEY.
XX
XX Abbott CA, Gorrell MD;
XX
XX WPI: 2002-454646/48.
XX
XX P-PSDB; AAE24169.
XX
XX
XX New dipeptidyl peptidase (DPP) peptidases, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
XX
XX Disclosure: Page 67-70; 91pp: English.
XX
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX polynucleotides encoding such proteins. The DPP peptidases are useful for
XX screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV (human immuno deficiency virus) infection. The present
XX sequence is an alternative version of murine DPP9 cDNA.
XX Note: This sequence is stated to be the same as that shown as
XX SEQ ID NO: 3 in figure 9 of the specification. However these sequences
XX differ.
XX
XX Sequence 3287 BP; 744 A; 970 C; 877 G; 696 T; 0 other;
Query Match 74.0%; Score 1936.4; DB 24; Length 3287;
Best Local Similarity 84.3%; Pred. No. 0;
Matches 2180; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 19 CACGGGACCCCAAGCGCCGACGAGGCGACGAGCCGACAGATGACCCGCGCCG 78
Db 27 CACGGGAGTCTCCCAAGTTGAGCAGTGGCCGACGGGAGCATGATGACCGGACAGC 86
QY 79 CTTCACAGTTCAGAGACACTGCTGGGACGGGCTCCGAGCATCATCCAGGCGGCCAA 138
Db 87 CTTCCTGTGTCAAGAGACTGCTGGGAGTGGGCTGGGAGATTTATCCAGGCGAGTGCA 146
QY 139 GTACTGGGCTTATTTGTCAACAAGGCGCCGACGACTTCCATTGTGTGCAAGACGA 198
Db 147 GTCTCGGGGCTTATTTGTCAACAAGGCGCCGACGACTTCCATTGTGTGCAAGACCTGA 206
QY 199 TGAAGTGGGCGCCGACCTCCGAGCCCTCTACTGCTGGGAGTGCATGATGAGCGCGAGA 258
Db 207 CGAGTCTGGCCCGCCGACCTCTACCGTCTCTATTACTGTGATGATGCGGACGCGGTGA 266
QY 259 GAACCTCCCTCTCTACTGCTGAGATTCACCAAGAGTCCGGAAGAGAGGCTGCTGCTCT 318
Db 267 GAACCTCCCTCTCTACTGCTGAGATTCACCAAGAGTCCGGAAGAGAGGCTGCTGCTGCT 326
QY 319 GTCTTGGAAGCAGATGCTGATATTTCAGGCGCAGCGCCGACCATGGGCTACTCTCG 378
Db 327 GTCTTGGAAGCAGATGCTGATGATTCAGGCGCAGCGCCGACCATGGGCTACTCTCGC 386
QY 379 GGAAGAGGAGCTGCTGAGGAGGGAAGAGCGGCGGCTGTTGGCATTCACCTCTACGA 438
Db 387 AGAGAGAGAGCTACTGCTGAGGAGGCGCAAGCGCTGGGCGTCTTGGAATCACTCTTATGA 446
QY 439 CTTCACAGGAGAGAGTGGCT 498
Db 447 CTTCACAGGAGAGAGTGGCT 506
QY 499 CGAGCGCGCAAGAGCGCTTCATGCTGCTCCCTATGAAACCGCTGGAATTCAGAGCCA 558
Db 507 GGAATGCTGCAAGAAATGCTTATGCTGCCGATGAACCCACTGGAAGATCAAGACTCA 566
QY 559 GTGCTCAGGGCGCCGAGTGGAGTGGACCCCAAAATCTGCGCCGACCCCTCTCTCTCTCT 618
Db 567 GTGCTCAGGGCGCCGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 626
QY 619 CATCAATTAACAGGACCTGTGGTGGGCAATGAGACAGGAGGAGGCGGCGCTGAC 678
Db 627 CATCAATTAACAGGACCTGTGGTGGGCAATGAGACAGGAGGAGGCGGCGCTGAC 686
QY 679 CTTCGCGCACCAAGTTCATCAATGCTCTGATGAGCCCAAGCTGCGGGGTGTGGCCAC 738
Db 687 CTTCGCGCACCAAGTTCATCAATGCTCTGATGAGCCCAAGCTGCGGGGTGTGGCCAC 746
QY 739 CTTCGCGCACCAAGTTCATCAATGCTCTGATGAGCCCAAGCTGCGGGGTGTGGCCAC 798
Db 747 CTTCGCGCACCAAGTTCATCAATGCTCTGATGAGCCCAAGCTGCGGGGTGTGGCCAC 806
QY 799 CTGGGAAGGTTTCAGAGGCGCTCAAGAGCTGCGAATCTCTATGAGGAGATGCTGAGTC 858
Db 807 TTGGGAAGGTTTCAGAGGCGCTCAAGAGCTGCGAATCTCTATGAGGAGATGCTGAGTC 866
QY 859 CGAGGTGAGGATTCATCACTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 918
Db 867 TGAAGTGAAGTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 926
QY 919 TCGGTACCCAGAGCAGAGGCGCAAGAAATCCCAAGATTGCTTGAATGCTGAGTTCGA 978
Db 927 CCCTTACCCAGAGCAGAGGCGCAAGAAATCCCAAGATTGCTTGAATGCTGAGTTCGA 986
QY 979 GACTGACAGCAGAGGCGCAAGAAATGCTGCTGAGCAGAGGAGAGAGAGTGTGACGCTTCAG 1038
Db 987 GACGAGACATCAGGCGCAAAATGCTGCTGAGCAGAGGAGAGAGAGTGTGACGCTTCAG 1046
QY 1039 CTGCTGTTCGCAAGGTGAGTACATGCGCAGGCGCGGTGGAACCCGAGTGGCAAAATA 1098
Db 1047 CTCCTTTTCCCAAAAGTGAATACATGCGCGGCTGCTGAGACAGGAGCGCAAAATA 1106

QY	1099	GCCTGGGCAATGTTTCTGTGGACGCGGCCCAAGACAGTGGCTCAGCTGTCTCTCTCCCCC	1158
Db	1107	TTGGCCGGCCCAATGTTTCTGTGGACGCTGCCAGCAACGGCTTTCAGCTGTCTCTCTCCCCC	1168
QY	1159	GGCCCTGTTTCATCCCGAGACAGGAATGAGGAGCAGCGGCAATAGCCTCTGCCAGAGCT	1218
Db	1167	TGCTCTCTTCAATCCCGGCCCTGTGAGATGTAGGCCCCAGCGGCAAGCAAGCTGCCAGAGCCGT	1228
QY	1219	CCCCGAGAAATGCCAGCCGATATGTGGTGTGACGAGAGTCCACCAAGCTGTGGATCAATGT	1278
Db	1227	CCCCAGAAATGTGACAGCCCTTTGTCACTATGAAAGAAAGTACCACCAATGTCTGGATCAAGT	1286
QY	1279	TCATGACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGAGCAAGCTCTGCTTTCGG	1338
Db	1287	CCAGACACATCTTCCACCCTGTTTCTCCAGGCTGAGGGCCAGAGCAAGCTTTGTGTTCTTCG	1346
QY	1339	CGCCCAATAATGCAAGACCGGCTCTGCCCCATTTGTGTAAGAAAGTACCGCCGTTTAAATNC	1398
Db	1347	TGCCAAGCAATGCAAGACTGGCTTCTGCAACCTGTACAGAGGTACACAGTGGAACTTAAAC	1406
QY	1399	CCAGGGCTACGATTTGAGTGGATGAGCCCTTTCAGCCCTCCGGGGGAGATGTAATTTAATGCCCAT	1458
Db	1407	CAAGGACTATGACGTGGAGCGGAACCCCTACACCCTACAGAAAGGTAGTTAATGGCCCAT	1466
QY	1459	TAAAGAGAGATTTCTCTTGACACAGCGGTGANTGGAGGTTTTGGCGAGGACGCGTCCAA	1518
Db	1467	CAAGAGAGAGGTCTCCCTTGACACAGTGGCGAGTGGGAGGCTTGTGGAGGACATGCGTCCAA	1526
QY	1519	GATCTGGGTCAATGAGAGACCAAGCGGTGTACTTCCAGGGCCACCAAGAACAGCCGCT	1578
Db	1527	GATCTGGGTCAAGAGAGAGCAAGAGCGGTGTACTTTCAGGTACCAAGAACAGCACCCGT	1586
QY	1579	GGAGACACCCTCTACCTTGGTTCAGACTATGAGGGCGCGGCGAGATGTAAGCTTCACAC	1638
Db	1587	GGAACTACCTCTATGTGGTCAAGCTACAGATCAGAGTACAGGCGAGATGTGCGGCTACAC	1646
QY	1639	GCCCGGCTTCCCATAGTGTCTTCATGAGCCAGAACTTGACATGTTGTGACGCCACTA	1698
Db	1647	GCTGGGCTTCCCAAGCTGTCTTCATGAGCCAGAACTTGACATGTTGTGAGCACTA	1706
QY	1699	CAGACGGTGGACGCGCGCGCGGCGACGCTCAAGAGCTGAGCGGCGCCGACAGCA	1758
Db	1707	CAGAGTGTGGACGCGCACCCCTGTGACATGTGTACAAAGCTGAGCGGCCCGAGATGA	1766
QY	1759	CCCCCTGCACAAGCAGCCCGCTTCTGGGTGATACATGATGAGGAGCAGCAGCTGCCCCC	1818
Db	1767	CCCACTGCACAAGCAACAGCGTTCTGGGGCANCAATGATGAGGCGCAATTTGCCCCC	1826
QY	1819	GGATTTATGTTCTCTCAGAGATCTTCCATTTCCACACGCGCTCGATGTGCGGCTCTACGG	1878
Db	1827	AGACTATGTGTGCCCCCTAGATCTTCCACTTCCACCCCTGCAGAGCTGACGCTCTACGG	1886
QY	1879	CATGATCTACAGGCCCCAGCGCTTGGACGCCAGGGAAGAGCAACCCACGCTCTCTTGT	1938
Db	1887	CATGATCTACAGGCACACACCCCTGCACCTTGGAGGAAGACACCCACGCTGTCTTGT	1946
QY	1939	ATATGAGAGCCCCCAGGTGAGCTGTGATATACCTCTTCAAGGATCAAGTACTTGG	1998
Db	1947	CTATGGGGGCCACAAGGTGAGTTGGTGAACAACACTCTTTAAGGCACTCAATACGTGCG	2006
QY	1999	GCTCAACACACTGGCTCTCCCTGGGCTACGCGCGTGTGTGATTTGACGGCGAGGCGCTCTG	2058
Db	2007	GCTAAATACACTGGCACTCTTGGGCTATGCTGTGGTGATCGATGTGCGGGGCTCTG	2066
QY	2059	TCACGAGAGGCTTCGTTCGAAGGGGCCCTGAAAAACCAATGGGCGCAGGTGGAGATCGA	2118
Db	2067	TCACCGGGGCTGTCACTTTCAGGGGGGCCCTGAAAAATCAAAATGGGCGCAGGTGGAGATTTGA	2126
QY	2119	GGACCAAGTGGAGGCGCTGTGATTTCTGGGCGGAGATGTGGCTTCATCGACCTGAGCGG	2178
Db	2127	GGACCAAGTGGAGGCGCTGTGATGACTGTGCTGTGAGAGTATGGCTTCATTTGACTTGAAGCG	2186
QY	2179	AGTTGCCATCATGCTGTGCTACAGGGGGCTTTCCTCTGCTCATAGGGGCTTAATCCAA	2238

Db	2187	AGTGGCCATTCATGGCTGTGCTCAACGGCGGCTTCCTCACTCATCATGAGGGCTCATCCAA	2246
Qy	2239	GGCCCGAGTGTTCAAAGGTGGCCATCGCGGGGTGCCCGGTACACGTCGTGGATGGCTTACGA	2298
Db	2247	GCCAACATGTGTTCAAGGTATGACCATGTGGGGCGCTCTGTCACTGTGTGGATGGCTTATGA	2306
Qy	2299	CACAGGCTACACTTAGCGCTACATATGACGTCCTTGAGAACACACGACGGCTATGAGGC	2358
Db	2307	CACAGGTACACGGGAACAGATACATGGATGTGTCCTCCGGAATTAACACGACGAAGCTATGAGGC	2366
Qy	2359	GAGGTTCGCTGGCCCTGTCACGTGGAGAAAGCTGCCCAATGAGCCCAACGCTTGCATTACT	2418
Db	2367	AGGGCTGTATGACCTCTGCATGTGGAGAAAGCTGCCCAATGAGCTTAACCGCGCTGCTTACTCT	2426
Qy	2419	CCAGGCGTCTCTGTGACCAAAACGTGCACCTTTTCCACACAAACTTCCTGCTGCCAACT	2478
Db	2427	CCACGGCTCTCTGTGACCAAGAACGTTCACTTCTTCCACAAATTTCTGTGGTGTCCAACT	2486
Qy	2479	GATCCGAGACGAGGAACCTTACAGGCTCCAGATCTACCCCAACGAGAGACACAGATTGCG	2538
Db	2487	GATCCGAGACGAGGAAGGCATACCAAGCTTACAGATCTACCCCAACGAGAGACATAGCATTCG	2546
Qy	2539	CTGCGCCGAGTCTGGGCGAGCACTATGAAATGACAGTGTGTCACCTTTCTACAGGATACCT	2598
Db	2547	CTGCGCGGAGTCTGGGAGACGATTACGAGGGTGAAGCTGTGCACCTTCTGCAAGAAACACT	2606
Qy	2599	CTGAGC 2604	
Db	2607	CTGAGC 2612	

	RESULT 15
ID	AAc75835
XX	AAc75835 standard; cDNA: 2461 BP.
XX	
AC	AAc75835;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human ORFX ORF1390 polynucleotide sequence SEQ ID NO:2779.
XX	
XX	
XX	Human; open reading frame; ORFX; detection: cytostatic; hepatotropic;
XX	vulnerary; antiparasitic; antiparkinsonian; neotropic; neuroprotective;
XX	anticonvulsant; osteopathic; antiarthritis; immunosuppressant; cardiant;
XX	immunosuppressant; thromolytic; coagulant; vasotropic; antidiabetic;
XX	hypotensive; dermatological; immunosuppressive; antinflammatory;
XX	antiviral; antibacterial; antifungal; antithematic; antithyroid;
XX	antianemic; gene therapy; cancer; proliferative disorder; hypertension
XX	neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX	cholesterol ester storage; systemic lupus erythematosus; infection;
XX	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX	bone damage; cartilage damage; antinflammatory disease; coagulation;
XX	thrombosis; contraceptive; ss.
OS	Homo sapiens.
XX	
PN	W0200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PE	31-MAR-2000; 2000MO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shinkets RA, Leach M;

XX MPI: 2000-602362/57.
DR P-PSDB: AAB41626.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 2024 -2026; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytosolic; hepatotropic; vlnetary;
CC antiporatic; antiparkinsonian; nootropic; immunoprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressive;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasoregic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antihemmatic;
CC antihypoid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal hemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 2461 BP: 574 A; 728 C; 691 G; 468 T; 0 other;

Query Match 71.4%; Score 1869.6; DB 21; Length 2461;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1909; Conservative 0; Mismatches 14; Indels 21; Gaps 1;

QY 11 ATGGCCACCCAGGCGGACCCCAAGCGCGAGGCGAGCGAGCGCCACAGATGACCG 70
DB 514 ATGGCCACCCAGGCGGACCCCAAGCGCGAGGCGAGCGAGCGCCACAGATGACCG 573
QY 71 GCCGCCGCTTCCAGGTGACAGACACTCTGTGGACGGGCTCCGAGCATCATCCAGCG 130
DB 574 GCCGCCGCTTCCAGGTGACAGACACTCTGTGGACGGGCTCCGAGCATCATCCAGCG 633
QY 131 AGCGGCAAGTACTGGGCTTATTGTCAACAAGCGCGCCACAGCTTCCAGTTGTGAC 190
DB 634 AGCGGCAAGTACTGGGCTTATTGTCAACAAGCGCGCCACAGCTTCCAGTTGTGAC 693
QY 191 AAGAGGATGATGTCGGGCGCCACTCCACGCGCTACTCTGCGGAGTCCCATNTAGC 250
DB 694 AAGAGGATGATGTCGGGCGCCACTCCACGCGCTACTCTGCGGAGTCCCATNTAGC 753
QY 251 AGCGGAGAGACTCCCTCTTACTCTGAGATTCACAGAGAGTCCGAAAGAGGCTTG 310
DB 754 AGCGGAGAGACTCCCTCTTACTCTGAGATTCACAGAGAGTCCGAAAGAGGCTTG 813
QY 311 CTGCTCTCTCTGAGAGCAGATGCTGATCATTTCCAGGCGCCGCCACCATGCGGTC 370
DB 814 CTGCTCTCTCTGAGAGCAGATGCTGATCATTTCCAGGCGCGCCACCATGCGGTC 873
QY 371 TACTCTGGGAGGAGAGCTGCTGAGGAGGAGGAAAGCGGCGGCGTTCGGCATCCAC 430
DB 874 TACTCTGGGAGGAGAGCTGCTGAGGAGGAGGAAAGCGGCGGCGTTCGGCATCCAC 933
QY 431 TCTACAGACTTCCACAGGAGAGTGGCTTCTTCTTCCAGAGGCGAGCAAGCCTCTTC 490
DB 934 TCTACAGACTTCCACAGGAGAGTGGCTTCTTCTTCCAGAGGCGAGCAAGCCTCTTC 993
QY 491 CACTGTGCGGACGGCGGCAAGAGCGCTTCATG-----GTGTC 529
DB 994 CACTGTGCGGACGGCGGCAAGAGCGCTTCATG-----GTGTC 1053

QY 530 CCTATGAACCGCTGGAAATCAAGACCCAGTGTACAGGCGCCGGATGAGCCCAAAATC 589
DB 1054 CCTATGAACCGCTGGAAATCAAGACCCAGTGTCTCAGGCGCCGGATGAGCCCAAAATC 1113
QY 590 TGCCCTGCGGACCGCTGCTTCTTCTTCTTCAATCAATCAAGAGCTTGTGGGCGCAC 649
DB 1114 TGCCCTGCGGACCGCTGCTTCTTCTTCTTCAATCAATCAAGAGCTTGTGGGCGCAC 1173
QY 650 ATCGAGACAGCGGAGAGGAGCGGCGTCACTTTCGCCACCAAGTTTATCCAAATGCTG 709
DB 1174 ATCGAGACAGCGGAGAGGAGCGGCGTCACTTTCGCCACCAAGTTTATCCAAATGCTG 1233
QY 710 GATGACCCCAAGTCTGCGGCGTGTGGCGCCCTTCGATCAACAGGAAGATGTGACCGCTTC 769
DB 1234 GATGACCCCAAGTCTGCGGCGTGTGGCGCCCTTCGATCAACAGGAAGATGTGACCGCTTC 1293
QY 770 ACTGGGTACTGTGGTGGCGCCACAGACCTTCGGAAGGTTGAGAGGCGTCAAGACGCTG 829
DB 1294 ACTGGGTACTGTGGTGGCGCCACAGACCTTCGGAAGGTTGAGAGGCGTCAAGACGCTG 1353
QY 830 CGAATCTGTATGAGAGAGTGAATGATCGAGGTGAGTGCATTCACGTCCCTCTTCT 889
DB 1354 CGAATCTGTATGAGAGAGTGAATGATCGAGGTGAGTGCATTCACGTCCCTCTTCT 1413
QY 890 GCGCTAGAAAGAAAGAAACGAGCTCTGATGCTACCCAGAGACAGCAGCAAAATGCC 949
DB 1414 GCGCTAGAAAGAAAGAAACGAGCTCTGATGCTACCCAGAGACAGCAGCAAAATGCC 1473
QY 950 AAGATTGCTTGAACCTGCTGAGTTCAGACTGACAGCCAGGCAAGATGCTTCGACC 1009
DB 1474 AAGATTGCTTGAACCTGCTGAGTTCAGACTGACAGCCAGGCAAGATGCTTCGACC 1533
QY 1010 CAGGAGAAAGAGCTGTGTGACGCTTCACTGCTGCTTCCGAAAGTGAATGATGCC 1069
DB 1534 CAGGAGAAAGAGCTGTGTGACGCTTCACTGCTGCTTCCGAAAGTGAATGATGCC 1593
QY 1070 AGGGCGGCTGAGACCGGGGATGGCAATAGCCTTGGGCAATGCTTGGACCGGCGCCAG 1129
DB 1594 AGGGCGGCTGAGACCGGGGATGGCAATAGCCTTGGGCAATGCTTGGACCGGCGCCAG 1653
QY 1130 CAGTGGCTCAGCTGCTGCTTCCCTCCCGGCGCTTTCATCCGAGCAGAGAAATGAG 1189
DB 1654 CAGTGGCTCAGCTGCTGCTTCCCTCCCGGCGCTTTCATCCGAGCAGAGAAATGAG 1713
QY 1190 GAGCAGCGGCTTACGCTTGTGCCAGAGCTTCCCAAGAAATGTCAGCCGTATGTGTGAC 1249
DB 1714 GAGCAGCGGCTTACGCTTGTGCCAGAGCTTCCCAAGAAATGTCAGCCGTATGTGTGAC 1773
QY 1250 GAGGAGTCCACCAACGTCTGATCAATGATGATCATCTTATCCCTTCCCAATCA 1309
DB 1774 GAGGAGTCCACCAACGTCTGATCAATGATGATGATCATCTTATCCCTTCCCAATCA 1833
QY 1310 GAGGAGAGAGCAGAGCTGCTTCTCCGCGCAATGATGATGATGATGATGATGATGAT 1369
DB 1834 GAGGAGAGAGCAGAGCTGCTTCTCCGCGCAATGATGATGATGATGATGATGATGATGAT 1893
QY 1370 TTGTACAAAGTACCGCGCTTTTAAATCCAGAGGCTTACGATGAGTGAAGCCTTCAGC 1429
DB 1894 TTGTACAAAGTACCGCGCTTTTAAATCCAGAGGCTTACGATGAGTGAAGCCTTCAGC 1953
QY 1430 CCGGCGGAGAGTGAATTTAAGTGGCCCATTAAGGAAGATTTGCTTACACAGGCGTAA 1489
DB 1954 CCGGCGGAGAGTGAATTTAAGTGGCCCATTAAGGAAGATTTGCTTACACAGGCGTAA 2013
QY 1490 TGGAGAGTTTGGCGAGGACAGGCTTCAAGATCTGGGCTCAATGAGAGACCAAGCTGCTG 1549
DB 2014 TGGAGAGTTTGGCGAGGACAGGCTTCAAGATCTGGGCTCAATGAGAGACCAAGCTGCTG 2073
QY 1550 TACTTCCAGGCGCACCAAGAGACGCGCTGAGACACACCTTACGTGTGATGATGAG 1609
DB 2074 TACTTCCAGGCGCACCAAGAGACGCGCTGAGACACACCTTACGTGTGATGATGAG 2133

```

QY 1610 GCGGCCGGCGAGATGTCAGGCTCACCACGCCGGCTTCTCCATAGCTGCTCCATGAGC 1669
    |||||||
Db 2134 GCGGCCGGCGAGATGTCAGGCTCACCACGCCGGCTTCTCCATAGCTGCTCCATGAGC 2193
    |||||||
QY 1670 CAGAACTTCGACATGTCGTACGCACTACAGCAGCGTGAAGCAGCGGCCCTGGTGAC 1729
    |||||||
Db 2194 CAGAACTTCGACATGTCGTACGCACTACAGCAGCGTGAAGCAGCGGCCCTGGTGAC 2253
    |||||||
QY 1730 GTCACAAAGCTGAGCGGCCCGAGCAGACCCCTGCACAAGCAGCCCGCTTCTGGGCT 1789
    |||||||
Db 2254 GTCACAAAGCTGAGCGGCCCGAGCAGACCCCTGCACAAGCAGCCCGCTTCTGGGCT 2313
    |||||||
QY 1790 AGCATGATGAGGAGCGCAGCTGCCCGGATTATGTTCTCCAGAGATCTTCATTTC 1849
    |||||||
Db 2314 AGCATGATGAGGAGCGCAGCTGCCCGGATTATGTTCTCCAGAGATCTTCATTTC 2373
    |||||||
QY 1850 CACACGGGCTCGGATGTCGGGCTCTACGGCATGATCAAGCCCGCAGGCTTGACAGCA 1909
    |||||||
Db 2374 CACACGGGCTCGGATGTCGGGCTCTACGGCATGATCAAGCCCGCAGGCTTGACAGCA 2433
    |||||||
QY 1910 GGGAGAAGCACCACCGTCTTC 1933
    |||
Db 2434 ATCACCAAAAATCTACGCTCTTC 2457
    |||

```

Search completed: December 12, 2002, 08:42:03
 job time : 400 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 12, 2002, 13:21:43 : Search time 74 Seconds

(without alignments)
4611.930 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646
Sequence: 1 MATTCPTADRGDAATDDP.....CPESGEHYETLLHPIQEYL 863

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 355320 segs, 19730502 residues

Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09976674/runat_04122002_162533_7241/app-query.fasta.1.1031
-DB=Published_Applications_NA -OFMT=fastap -SUFFIX=rpnb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -MATRIX=blcsunm2
-TRANS-human40.csl -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPPISIZE=500 -MINLEN=0
-MAJLEN=200000000 -USER=US09976674@cgn2_1.1.33 -runat_04122002_162533_7241
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6
-YGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEM_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4646	100.0	2617	US-09-976-674-4	Sequence 4, Appl
2	4646	100.0	4219	US-09-976-674-28	Sequence 28, Appl
3	4646	100.0	4302	US-09-976-674-24	Sequence 24, Appl
4	4558.5	98.1	4180	US-09-976-674-36	Sequence 36, Appl

5	4558.5	98.1	4263	10	US-09-976-674-34	Sequence 34, Appl
6	4337	93.3	4076	10 <th>US-09-976-674-32</th> <th>Sequence 32, Appl</th>	US-09-976-674-32	Sequence 32, Appl
7	4337	93.3	4159	10 <th>US-09-976-674-30</th> <th>Sequence 30, Appl</th>	US-09-976-674-30	Sequence 30, Appl
8	4249.5	91.5	4037	10 <th>US-09-976-674-40</th> <th>Sequence 40, Appl</th>	US-09-976-674-40	Sequence 40, Appl
9	4249.5	91.5	4120	10 <th>US-09-976-674-38</th> <th>Sequence 38, Appl</th>	US-09-976-674-38	Sequence 38, Appl
10	2870	61.8	2671	10 <th>US-09-976-674-12</th> <th>Sequence 2, Appl</th>	US-09-976-674-12	Sequence 2, Appl
11	2850	61.3	4829	10 <th>US-09-976-674-12</th> <th>Sequence 12, Appl</th>	US-09-976-674-12	Sequence 12, Appl
12	2642.5	56.9	4685	10 <th>US-09-976-674-22</th> <th>Sequence 22, Appl</th>	US-09-976-674-22	Sequence 22, Appl
13	2608	56.1	4676	10 <th>US-09-976-674-20</th> <th>Sequence 20, Appl</th>	US-09-976-674-20	Sequence 20, Appl
14	2576	55.4	2411	10 <th>US-09-976-674-26</th> <th>Sequence 26, Appl</th>	US-09-976-674-26	Sequence 26, Appl
15	2548	54.8	2079	12 <th>US-10-044-090-843</th> <th>Sequence 843, Appl</th>	US-10-044-090-843	Sequence 843, Appl
16	2388.5	51.4	4523	10 <th>US-09-976-674-8</th> <th>Sequence 8, Appl</th>	US-09-976-674-8	Sequence 8, Appl
17	2204.5	47.4	4309	10 <th>US-09-976-674-14</th> <th>Sequence 14, Appl</th>	US-09-976-674-14	Sequence 14, Appl
18	647	13.9	1356	10 <th>US-09-976-674-3</th> <th>Sequence 10, Appl</th>	US-09-976-674-3	Sequence 10, Appl
19	498.5	10.7	4835	10 <th>US-09-917-800A-1570</th> <th>Sequence 1570, Appl</th>	US-09-917-800A-1570	Sequence 1570, Appl
20	496	10.7	832	10 <th>US-09-976-674-18</th> <th>Sequence 18, Appl</th>	US-09-976-674-18	Sequence 18, Appl
21	487.5	10.5	3407	12 <th>US-10-002-593-5</th> <th>Sequence 5, Appl</th>	US-10-002-593-5	Sequence 5, Appl
22	487	10.5	281	10 <th>US-09-967-550-987</th> <th>Sequence 987, Appl</th>	US-09-967-550-987	Sequence 987, Appl
23	444	9.6	2583	10 <th>US-09-976-674-6</th> <th>Sequence 6, Appl</th>	US-09-976-674-6	Sequence 6, Appl
24	444	9.6	4541	10 <th>US-09-976-674-42</th> <th>Sequence 42, Appl</th>	US-09-976-674-42	Sequence 42, Appl
25	426	9.2	2814	9 <th>US-09-954-531-367</th> <th>Sequence 367, Appl</th>	US-09-954-531-367	Sequence 367, Appl
26	426	9.2	2814	10 <th>US-09-962-832-108</th> <th>Sequence 108, Appl</th>	US-09-962-832-108	Sequence 108, Appl
27	426	9.2	2814	10 <th>US-09-954-456-1148</th> <th>Sequence 1148, Appl</th>	US-09-954-456-1148	Sequence 1148, Appl
28	426	9.2	2815	10 <th>US-09-965-606-1</th> <th>Sequence 1, Appl</th>	US-09-965-606-1	Sequence 1, Appl
29	411.5	8.9	4496	10 <th>US-09-976-674-44</th> <th>Sequence 44, Appl</th>	US-09-976-674-44	Sequence 44, Appl
30	266.5	5.7	535	10 <th>US-09-604-87A-428</th> <th>Sequence 428, Appl</th>	US-09-604-87A-428	Sequence 428, Appl
31	266.5	5.7	535	12 <th>US-10-007-805-428</th> <th>Sequence 428, Appl</th>	US-10-007-805-428	Sequence 428, Appl
32	249	5.4	1958	10 <th>US-09-974-300-224</th> <th>Sequence 224, Appl</th>	US-09-974-300-224	Sequence 224, Appl
33	216.5	4.7	2383	12 <th>US-10-044-090-521</th> <th>Sequence 521, Appl</th>	US-10-044-090-521	Sequence 521, Appl
34	193	4.2	497	10 <th>US-09-919-580-244</th> <th>Sequence 244, Appl</th>	US-09-919-580-244	Sequence 244, Appl
35	183	3.9	539	9 <th>US-10-046-935-2043</th> <th>Sequence 2043, Appl</th>	US-10-046-935-2043	Sequence 2043, Appl
36	183	3.9	539	9 <th>US-09-878-178-2043</th> <th>Sequence 2043, Appl</th>	US-09-878-178-2043	Sequence 2043, Appl
37	181	3.9	529	9 <th>US-10-046-935-2021</th> <th>Sequence 2021, Appl</th>	US-10-046-935-2021	Sequence 2021, Appl
38	181	3.9	529	9 <th>US-09-878-178-2021</th> <th>Sequence 2021, Appl</th>	US-09-878-178-2021	Sequence 2021, Appl
39	168.5	3.6	244	10 <th>US-09-960-352-184</th> <th>Sequence 184, Appl</th>	US-09-960-352-184	Sequence 184, Appl
40	166	3.6	620	10 <th>US-09-976-674-16</th> <th>Sequence 16, Appl</th>	US-09-976-674-16	Sequence 16, Appl
41	156.5	3.4	1797	10 <th>US-09-974-300-338</th> <th>Sequence 338, Appl</th>	US-09-974-300-338	Sequence 338, Appl
42	153	3.3	1869	10 <th>US-09-966-803-1</th> <th>Sequence 1, Appl</th>	US-09-966-803-1	Sequence 1, Appl
43	142	3.1	312	10 <th>US-09-728-445-436</th> <th>Sequence 436, Appl</th>	US-09-728-445-436	Sequence 436, Appl
44	118.5	2.6	2508	10 <th>US-09-934-223-3</th> <th>Sequence 3, Appl</th>	US-09-934-223-3	Sequence 3, Appl
45	118.5	2.6	4667	10 <th>US-09-934-223-1</th> <th>Sequence 1, Appl</th>	US-09-934-223-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-976-674-4
: Sequence 4, Application US/09976674
: Patent No. US20020115843A1
GENERAL INFORMATION:
: APPLICANT: Ol, Steve
: APPLICANT: Akinsanya, Karen
: APPLICANT: Riviere, Pierre
: APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 2617
TYPE: DNA
ORGANISM: Homo sapiens
US-09-976-674-4

Alignment Scores:
Pred. No.: 0
Score: 4646.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%

Length: 2617
Matches: 863
Conservative: 0
Mismatches: 0

Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

QY	1	MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspAspPro	20
Db	11	ATGGCCACCCACGGGGACCCCAAGCGCCGACCGAGGGAGCGACACCCCCACATACACCG	70
QY	21	AlaAlaThrPheGlnValGlnLysHisSerTrpAspGlyLeuAspSerLeuLeuHisGly	40
Db	71	GCCGCCCGCTCCAGGTGCGAAGAACACTCGTGGGACGGGCTCCGAGACTCATCCAGGC	130
QY	41	SerArgLysTrpSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln	60
Db	131	AGCCGCAAGTACTCGGGCCCTCATTTGTCACAAAGGCCCCACAGACTTCAGTTGTGACG	190
QY	61	LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGluMetProTyrGly	80
Db	191	AAGACGATGATGTGGGCCCACTCCACCGGCTCTACTACCTGGAAATGCCAATATGGC	250
QY	81	SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgGlyGluAlaLeu	100
Db	251	AGCCGAGAGAACTCCCTCCTACTCTGAGATTCCCAAGAGTCCGGAAGAAGCTCTG	310
QY	101	LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal	120
Db	311	CTGCTCCGTGCTCGAAGCAGATGCTGATCATTTCCAGGCCACGCCCCATCATGGGGTC	370
QY	121	TyrSerArgGluGluGluLeuAlaArgGluValArgLysArgLeuGlyValPheGlyLeuThr	140
Db	371	TACTCTCGGGAGGGAGGAGCTGTGAGGAGACGGAAAGCCTGGGGTCTTCCGCAATCC	430
QY	141	SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe	160
Db	431	TCCTACACACTCCACACGCGAGAGGGGCTTCTCTCTCCAGGCCAGCAACGCTCTTC	490
QY	161	HisGlyAsnArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle	180
Db	491	CACGTGTGCGACGGCGGCAAGAACGGCTTCATGTGTCCCATAGAAACCGTGGAAATC	550
QY	181	LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe	200
Db	551	AAGACCCAGTGTCTAGGGCCCCGGATGAGACCCAAATATGTGCCCTGCGACCTCTTC	610
QY	201	PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg	220
Db	611	TTTCCTCTTCATCAATAAACACGACGACTGTGGGTGGCCAAATCATGACAGCAGGACGG	670
QY	221	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerLeuGly	240
Db	671	CGGTGTACCTTGTGCCACCAAGGTTTATTCAAATGCTCTGGAATGCCCAAGTCTCGGGT	730
QY	241	ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpTrpCysPro	260
Db	731	GTGGCCACCTTCGTCATACAGGAAGATTGACCGCTTACAGGGGATCGTGGTGGCCCC	790
QY	261	ThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluVal	280
Db	791	ACACCCCTCCGGGAAGGTTTCAGAGGGCTTCAGAACCTTCGAAATCTGATAGGAATCC	850
QY	281	AspGluSerGluValGluValIleLeuHisValProSerProAlaLeuGluGluArgLysThr	300
Db	851	GATGAGTCCAGGTGGAGGTCAATTACAGCTCCCTCTCTGCGCTTGAAAGAAAGGAGACG	910
QY	301	AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla	320
Db	911	GACTCGATCGGTAACCCACAGACAGCAGCAAGAATCCCAATTTGCCTTGAACCTGGCT	970
QY	321	GluPheGlnThrAspSerGlnGlyLysIleValSerThrGluGluGluLeuValGln	340
Db	971	GAGTTCACAGACTACACGCCAGGGCAAGATCGTCTGACCCAGAGAAAGAGCTGTGTGAG	1030

Qy	341	ProPheSerSerLeuPheProLysValGluTyrIleAlaIArgAlaGlyTrpThrArgAsp	360
Db	1031	CCCTTCAGCTGCTGCTCCGGAAGGTGAGATACATGCGCAGGCGGGGTGACCCGGGAT	1090
Qy	361	GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu	380
Db	1091	GGCAAAATACGGCTGGGCCAATGTTCTGGACGGGCCCCACACTGGCTCCAGCTCTCTCC	1150
Qy	381	LeuProProAlaLeuPheIleProSerThrGlnAsnGluGlnArgLeuAlaSerAla	400
Db	1151	CTCCCCCGGCCCTGTTCAATCCCGAGCACAGAAATGAGAGACAGCGGTACCTCTCC	1210
Qy	401	ArgAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTrp	420
Db	1211	AGACCTGTGCCACAGAAATGTCCAGCGGTATGTGTGTACAGAGAGTCCACCAAGCTTGG	1270
Qy	421	IleAsnValHisAspIlePheTyrProPheProGlnSerGluGlyGluAspLeuLys	440
Db	1271	ATCAATGTTCATGACATCTTCATCCCTTCCCAATCAGAGGGAGAGACAGCTCTGC	1330
Qy	441	PheLeuArgAlaAsnGluCysLysTrpGlnPheCysHisLeuTyrLysValThrAlaAla	460
Db	1331	TTTTCTCGCGCCAAATGAAATCAAGACCGGCTTCTGCAATTTGTACAAAGTCCAGCCGTT	1390
Qy	461	LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGlyGluAspLeuPheLys	480
Db	1391	TTAAATATCCAGGGCTACGATGTGAGTGAAGCCCTTCAGCCCGGGGAAGATGAATTTAAG	1450
Qy	481	CysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArgHis	500
Db	1451	TGCCCATTTAAGGAAGAAATGCTCTGACACAGGGGAATGGGAAGTTTGGCGAGGAC	1510
Qy	501	GlySerLysIleTrpValAsnGluGluTrpTrpLysLeuValTyrPheGluGlnTyrHisLysP	520
Db	1511	GGCTCCAGATCTGGGTCATGAGGAGACCAAGCTGTGTACTTCCAGGACCCAGACAGAC	1570
Qy	521	ThrProLeuGlnHisHisLeuTyrValSerTyrGluAlaAlaGlyGluIleValArg	540
Db	1571	ACGGCGCTGAGACACCCCTCTACGTGGTACAGTATGAGCGCGGAGATCGTACGC	1630
Qy	541	LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheAla	560
Db	1631	CTCACACAGCCCGGCTTCTCCATTAACCTCTCCATGAGGACCAACTGGATGTTGTCGTC	1690
Qy	561	SerHisTyrSerSerValSerThrProCysValHisValTyrLysLeuSerGlyPro	580
Db	1691	AGGCACATACAGCAGCGTGAACAGCCGCCCTCGCTGCACGTATCAACCTGACGGCGCCC	1750
Qy	581	AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaAlaSer	600
Db	1751	GAGCAGACACCCCTGCGACAGCAGCCCGCTTCTGGGCTAGCATGATGAGGACACCGAC	1810
Qy	601	CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg	620
Db	1811	TGCCCCCGGAATTAATGTTCTCCAGAGATCTTCCATTTCCACACCGCTCGGATGTGGG	1870
Qy	621	LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal	640
Db	1871	CTCTACGGCATGATCTACAAAGCCCCACGGCTTCCAGCACCGGAAAGAACACCCACGCTC	1930
Qy	641	LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys	660
Db	1931	CTCTTTTATATATGAGAGCCCCCGAGGTCAGCTGGTGTAATTAATCTCTTCAAGAGCATCAAG	1990
Qy	661	TyrIleLeuArgLeuAsnThrIleuAlaSerLeuGlyTyrAlaValValValIleAspGlyArg	680
Db	1991	TACTTGGGGCTCAACACACAGCGCTCCCTCGGGCTCAGCGCTGGTGTATTTAGCGGACAG	2050
Qy	681	GlySerCysGlnArgGlyLeuArgPheGluGluValAlaLeuLysAsnGlnMetGlyGlnVal	700
Db	2051	GGCTCTCTGTACGCGAGGGCTTCCGTTCTGAAAGGGAGCCCTGAAAAACCAATGGGACGGTG	2110
Qy	701	GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp	720

Db	427	GC	CCCCCGCCTTCCAGGTGCGAAAGCACTGCTGGGAGGGGCTCCGAGCATCCACGGC	486
Qy	41	Ser	ArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln	60
Db	487	AG	CGCAAGTACTCGGGCTCATTTGTCACAAAGAGGGCCCCACGACTTCAGATTGTGTGAG	546
Qy	61	Lys	ThrAspGlnSerGlyProHisSerHisAlaGluTyrTyrLeuGlyMetProTyrGly	80
Db	547	AAG	CGGATGAGTCTGGGGCCCCACTCCACCGGCTCTACTACTGGGAATCCATATGCC	606
Qy	81	Ser	ArgGlnAsnSerLeuLeuTyrSerGluIleProLysValArgLysGlnAlaLeu	100
Db	607	AG	CGGAGNAGACTCCCTCTTACTGTGAGATTTCCAGAGAGTCCGGAAAGAGGCTCTG	666
Qy	101	Leu	LeuLeuSerTyrLysGlnMetLeuAsnHisPheGlnAlaThrProHisIleGlyVal	120
Db	667	CTG	CTCGTCTCTGGAGGAGATGTGATCATTTCCAGGCACGCCACCATGGGGTC	726
Qy	121	Tyr	SerArgGlnGlnGluLeuLeuArgGlnArgLysArgLeuGlyValPheGlyIleThr	140
Db	727	TACT	CTCGGGAGGAGCTGTAGAGGACGGAAACGGCTGGGGCTCTCGCATCCACC	786
Qy	141	Ser	TyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe	160
Db	787	TCCT	ACGACTTCCACACGAGAGTGGCTCTTCTCTTCCAGGCCACGACACAGCTTCTTC	846
Qy	161	His	CysArgAspGlnGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle	180
Db	847	CAC	TGCGCGCGACGGCGCAGAAACGGCTTCATGTGTCTCCCATATGAACCGCTGGAATTC	906
Qy	181	Lys	ThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe	200
Db	907	AAG	ACCAAGTCTCTAGGGCCCCCGGATGTAGACCCCAAAATGTGCCCTTCGACACCTCTCTTC	966

Qy	201	PheSerPheIleAsnAsnSerAspIleThrPvalAlaAsnIleGluThrGlyIleIuIuArg	220
Db	967	TTTCCTCTTCAATCAATAACAGGACCTGTGGTGCCAACTAGAGACAGGACGAGGCG	1028
Qy	221	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValIleuAspAspProIysSerIacIy	240
Db	1027	CGGTGACCTTCTCCACCAAGGTTATCCAAATGCTCTGGATGAGCCCCCAAGTCTGGGGT	1086
Qy	241	ValAlaThrPheValIleGlnGluIuPheAspArgPheThrGlyTyrTrpTrpCysPro	260
Db	1087	GTGGCACCTTCGTCAATACAGGAAGGTTGACCCCTTCACTGGGAGTACGTGGTGGCCCC	1146
Qy	261	ThrAlaSerTrpGluIysArgIuIuIysThrIleuArgIleIeuThyGluIuVal	280
Db	1147	ACAGCTCTCTGGGAAGGTTCAGAGGGCTTAAACCCCTCGAATCTGTATAGGAAGTTC	1206
Qy	281	AspGluSerGluValGluValIleHisValProSerProAlaLeuGluIuIuArgIysThr	300
Db	1207	GATGAGTCCGAGGGGAGAGTCATTACAGTCCCTCTCTGCGCTAACAAGAAGAGACG	1266
Qy	301	AspSerTyrArgTyrProArgThrGlySerIysAsnProIysIleAlaIeuIysLeuAla	320

Qy	321	GIuPhEgInTrRAspSerGInGlyIleValSerThrGluGluGlySGluGluValGln	340
Db	1327	GAGTTCAGACTGCAGCGCAAGCAATGCTCTCGACCAGAGCAAGAGCGTGGTGCAG	1386
Qy	341	ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGluTyrTrpThrArgAsp	360
Db	1387	CCCTTCACACTCGCTTCCGAAAGTGCATACATCGCCAGGGCCGGGTGGAGACCCGGAGT	1446

DB 1447 GGCACATACGGCTGGGCACTTTCCTGACGGGCCCAAGAGTGGCTCAGCTCCTCTC 1506

```
Db 1507 CTCCCCCGCCCTGTCATCCGACAGATGAGGACAGCGGCTCTGCC 1566
Qy 401 ArgAlaValProArgAsnValGlnProTyrValTyrGluGluValThrAsnValTyr 420
Db 1567 AGAGCTGTCCCGACAGATGTCACGCCGATGTGGTACGAGGAGGTCCACCAACGTCGCG 1626
Qy 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluAspGluLeuGly 440
Db 1627 ATCATATGTCATGACATCTCTATCCCTCCCAATCAGAGGAGAGGACGACCTCTGCC 1686
Qy 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460
Db 1687 TTTCTCCGCCCAATGATGATGACAGACGCGCTTGCGCATTTGTACAAAGTCAACGCGCTT 1746
Qy 461 LeuLysSerGlnGlyTyrAspTyrSerSerGluProPheSerProGluGluAspGluPheLys 480
Db 1747 TTTAAATCCCGAGGCTACGATGTGAGTGAAGCCCTTCAGCCCGGGGAGAAATGATTTTAA 1806
Qy 481 CysProIleLysGluGluIleAlaLeuThrSerGlyLysValLeuAlaArgHis 500
Db 1807 TGCCCCATTAAAGAAAGATTGCTCTGACACAGCGGTGATGGAGGTTTGGCAGAGCAC 1866
Qy 501 GlySerLysIleTyrPheValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp 520
Db 1867 GGCCTCCAAAGATCTGGCTCAATGAGAGACCAAGCTGGGTGTACTTCCAGGGACCAAGGAC 1926
Qy 521 ThrProLeuGlnHisHisLeuTyrValValSerTyrGluAlaAlaGluIleValArg 540
Db 1927 AGCGCGGTGGAGACCAACCTCTACGTGGTCAATGAGCGCGGCGGAGATGTTAGCC 1986
Qy 541 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
Db 1987 CTCACCAAGCGCGGCTCTCCATGAGCTGCTCCATGAGCCGAACCTTCGACATGTCGTC 2046
Qy 561 SerHisTyrSerSerValSerThrProCysValHisValTyrLysLeuSerGlyPro 580
Db 2047 AGCCACTACACACACCGTAGAGACGCGCGCTGCTGCACGCTACCAAGCTAGAGGCGCCC 2106
Qy 581 AspAspAspProLeuHisLysGlnProArgPheThrAlaSerMetMetGluAlaAlaSer 600
Db 2107 GACGACGACCGCCCTGCACAAACGACCGCGCTTGGGTGTACATGATGAGGACGCGCAG 2166
Qy 601 CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 620
Db 2167 TGCCCCCGGTTATGTTCTCCACAGAGATCTTCATTTCCACACGCGCTCGATGTGCGG 2226
Qy 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysHisProThrVal 640
Db 2227 CTCTACGGCATGATCTACAAAGCCCGCCCTTGACAGCCAGGAGAAAGACCCCAACGCTC 2286
Qy 641 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
Db 2287 CTCTTTGTATATGAGAGCGCCCGACAGTGCAGTGTGAATAACTCTTCAAAAGCATCAG 2346
Qy 661 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArg 680
Db 2347 TACTTGGGGCTCAACACACTGCGCTCCCTGGGCTACGCGGGTGTGATGAGCGCAGG 2406
Qy 681 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGluVal 700
Db 2407 GGCCTCTGTCAAGGAGGCTTCGCTTGAGAGGCGCCCTGAAAACCAACAAATGGGACAGTG 2466
Qy 701 GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
Db 2467 GAGATGACAGGACAGGAGGCGCTGACAGTGTGGCCGAGAGATGTGGTTATATGAC 2526
Qy 721 LeuSerArgValAlaIleHisGlyTyrPheTyrGlyLysPheLeuSerLeuMetGlyLeu 740
Db 2527 CTGAGCGCAATGTCATTCATGCTGCTACAGGGGCGCTTCCCTGCTCATATGGGCTA 2586
Qy 741 IleHisTyrProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTyrMet 760
Db 2587 ATCCACAAGCGCCAGGTTTCAGAGTGCATCCGCGGGTGCCCGGTACCGCTGTGATG 2646
```

```
Qy 761 AlaTyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGly 780
Db 2647 GCCTACGACACAGGCTACACTGACGCTACATGAGACGCTCCCTGAGAAACACACAGCGCC 2706
Qy 781 TyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeu 800
Db 2707 TATAGCGGGGTTCCGCGGCGCTCCAGCTGTGAGAAAGTGGCCCAATGAGCCCAACCGCTTG 2766
Qy 801 LeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVal 820
Db 2767 CTATTCCTCCAGCGCTTCCTGTGACGAAACGTGCACATTTTCCACAAACTTCTCTGTC 2826
Qy 821 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHis 840
Db 2827 TCCCACTGATCCGAGCAGGAGGAAACCTTACAGCTCCAGATCTACCCCAACAGAGACAC 2886
Qy 841 SerIleArgCysProGluSerGlyGluHisTyrGluValThrLeuLeuHisPheLeuGln 860
Db 2887 AGTATTCGCTGCGCCGAGTGGGCGAGCACATATGAACTGTGACATTTGCTACAG 2946
Qy 861 GluTyrLeu 863
Db 2947 GAATACCTC 2955

RESULT 3
US-09-976-674-24
; Sequence 24, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPTV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 4302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-24

Alignment Scores:
Pred. No.: 0 Length: 4302
Score: 4646.00 Matches: 863
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-976-674-3 (1-863) x US-09-976-674-24 (1-4302)
Qy 1 MetAlaThrThrGlyThrProThrAlaAspArgLysAspAlaIleAlaThrAspAspPro 20
Db 367 ATGGCCACCAACCGGACCCCAACAGCGCGAGCCGACGCGACGCGGACGATGAGACCGG 426
Qy 21 AlaAlaArgPheGlnValGlnLysHisSerThrAspGlyLeuArgSerIleIleHisGly 40
Db 427 GCGCGCGGCTTCAGAGTGCAGAAAGCACTCGTGGAGCGGCTCCGAGACATTCACAGGC 486
Qy 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 487 AGCGCAAGTACGCGGCTCATTTGTCAACAAGCGCCCAACAGCACTTCAGTTGTGACAG 546
Qy 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
Db 547 AAGAGGATGAGTCTGGGCGCCCACTCCACCGGCTTACTGACCTGGAAATGCCATATGGC 606
```

OY	81	SerArgGluAsnSerLeuTyrSerGluIleProLysValArgLysGluAlaLeu	100
Db	607	AGCCGAGAGAACTCCCTCTACTCTGAGATTCCCAAGAGCTCGGAAAGAGCTCTG	666
OY	101	LeuLeuLeuSerTyrPylsGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyAla	120
Db	667	CTGCTCTCTCTGGAAAGCAGATGCGATCATTTTCCAGGCCACGCCCTTACATGGGCTC	726
OY	121	TyrSerArgGluGlnGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyTlIethr	140
Db	727	TACTCTGGGAGAGAGAGCTGCTGAGGAGCGGAAACGCTTGGGGCTTCTGGCATCAC	786
OY	141	SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe	160
Db	787	TCTTACACTCTCCACACGCGAGTGGCCCTTCTCTTCCAGGCGACAGAACGCTCTTC	846
OY	161	HisGlyAsnArgspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle	180
Db	847	CACGCGCGAGCGCGGCAAGAAAGCGCTTCATGAGTGTCCTCATTAAGAACCGCTGGAATC	906
OY	181	LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe	200
Db	907	AAGACCCAGTCTCTAGGGCCCCCGGATGAGACCCCAAAATGTGCCCTCCGACCTCTCTC	966
OY	201	PheSerPheIleAsnAsnSerAsnLeuTyrPvalAlaAsnIleGluThrGlyGluGluArg	220
Db	967	TTTCTCTTCATCATATAACACGACGCTGTGGTGCCCAATCGAAGACGCGAGAGCGG	1026
OY	221	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAsnAspProLysSerAlaGly	240
Db	1027	CGGCTGACCTTCTCCACCAAGGTTTATCATATGCTCTGAGATGCCCAAGCTCTCGGGT	1086
OY	241	ValAlaThrPheValIleGlnGlnGluPheAspArgPheThrGlyTyrTyrProCysPro	260
Db	1087	GTGGCCACCTTCGTCATACAGAAAGATTGCAACCGCTTCACTGGGTACTGGGGTCCCCC	1146
OY	261	ThrAlaSerTyrPglGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGluGluVal	280
Db	1147	ACACCCCTCTCGGGAAGGTTGAGAAGGCGCTCAACACCTCGAATCTGTATAGAGAAATC	1206
OY	281	AspGluSerGluValGluValIleHisValProSerProAlaLeuGlnAlaArgLysThr	300
Db	1207	GATGATGCCAGGTTGGAGGTCATTCACTCCCTCTCTCGCCTTGAAAGAAAGGACGACG	1266
OY	301	AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla	320
Db	1267	GACTCGTATCGGTACCCACAGACAGCAGAGAAGAAATCCCAAGATTGCTTGAACCTGCT	1326
OY	321	GluIleGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln	340
Db	1327	GAGTTCACAGACTGCACCCAGGCGCAAGATCGTCTCACCCAGGAAGAGACCTGGTGCAG	1386
OY	341	ProPheSerSerLeuPheProLysValGlyTyrIleAlaArgAlaGlyTyrThrArgAsp	360
Db	1387	CCCTTCAGCTGCGTTCGCCGAAGGTGGAGTACATGCCACAGGCGCGGGTGGACCCGGCAT	1446
OY	361	GlyLysTyrAlaTyrPalametPheLeuAspArgProGlnIleTyrLeuGlnIleuValLeu	380
Db	1447	GGCAAAATACGCTGGCGCATGTTCTGAGACGGGCCCCACACTGGCTCCAGCTGCTCTC	1506
OY	381	LeuProProAlaLeuPheIleProSerThrGluAsnGluGluGluArgLeuAlaSerAla	400
Db	1507	CTCCCCCGGCGCTGTTCATCCCGAGCACGAGAAATGAGAGACAGCGCTGACCTCTGCC	1566
OY	401	ArgAlaValProArgAsnValGlnProTyrValValTyrGlnGluValThrAsnValTyrP	420
Db	1567	AGAGCTGTCCCCAGAGATGTCCAGCGGTATGTGTGTACGAGAGAGTCAACCACTCTGG	1626
OY	421	IleAsnValHisAspIleIethrTyrProPheProGlnSerGluGlyGluAspGluLeuCys	440
Db	1627	ATCATATGTTTCATGACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGAGACGACTCTGC	1686

QY	441	PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal	460
Db	1687	TTTCTCCGGCGCATGAAATCAAGACCGGGCTTCGCAATTTGTACAAAGTACACCGCCCTT	1746
QY	461	LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGluAsnGluPheLys	480
Db	1747	TTTAAATCCCAAGGGCTACGATGTGAAGTAGACCTTCATGCCCCGGGGGAAGATGAATTAAAG	1806
QY	481	CysProIleuLysGlnIleAlaLeuThrSerGlyLysTrpGluValIleuAlaArgHis	500
Db	1807	TGCCCATTTAAGAGAAAGATGTGCTCGACACGGGTGATGGAGAGTTTGGCAGAGCAC	1866
QY	501	GlySerLysIleTrpValAsnGluLysLysLeuValTyrPheGlnGlyThrLysAsp	520
Db	1867	GGCTCCAAAGATCTGGGTCCATGTAGAGACCAACCTGGTACTTCCAGGCGACCAAGAC	1926
QY	521	ThrProLeuGlnHisIleLeuTyrValValSerTyrGluAlaIleGlyLysIleValArg	540
Db	1927	ACGGCGCTGGAGACCCACCTCTACGTGGTACGTAAAGAGCGGGCGGAGATCTGTACG	1986
QY	541	LeuThrThrProGlyPheSerHisSerCysSerMetSerGluAsnPheAspPheVal	560
Db	1987	CTCACACAGCCCGGGCTTCTCCCATAGCTGCTCATATGACCAAACTTGACATGTGGTGC	2046
QY	561	SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro	580
Db	2047	AGCCACTACACACACCGTACACGCGCCCGCTCGCTGCATGCTACAGCTGACGGCGCCC	2106
QY	581	AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaIleSer	600
Db	2107	GACGACGACCCCTGACCAAGACGACCCCGCTTGGGGTAGATGATGAGGACACCGACG	2166
QY	601	CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg	620
Db	2167	TGCCCGCGGATTTATGTTCTCTCCAGAGATTTTCATTCCACACGCGCTCGATGTGGGG	2226
QY	621	LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysHisProThrVal	640
Db	2227	CTTACGCGCATGATCTCAAGACCCCAAGCGCTTGCACCCAGGAGAAGACCCCAACGCTC	2286
QY	641	LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyLys	660
Db	2287	CTCTTGTATATGAGAGGCCCCACAGGTGCACATGTGTAATCTCTTCAAAAGCATCAAG	2346
QY	661	TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArg	680
Db	2347	TACTTGGCGCTCAACACACTGGCTCCCTTGGCTACGCCGTGTGTATACGGCACGG	2406
QY	681	GlySerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyLysVal	700
Db	2407	GGCTCTGTTCAGAGAGGCTTCGGTTCTGAAAGGGGCCCTGAATAAACCAATGGGCCAGGTG	2466
QY	701	GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp	720
Db	2467	GAGATCGAGGACCGAGGTGGAGGGCGCTGCATTTGTGGCCGAAGAAGTATGGCTTCATCGAC	2526
QY	721	LeuSerArgValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeu	740
Db	2527	CTGACCGAGTGGCATTCATGGCTGTCTACGGGGGCTTCTCTCTGCTCATGGGGCTA	2586
QY	741	IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMet	760
Db	2587	ATCCACAAGCCCCAGGTGTTCAAGAGTGGCCATCGCGGAGGCCCGGTACCGTGTGATG	2646
QY	761	AlaTyrAspThrGlyTyrThrGlnArgTyrMetLeuAspValProGluAsnAsnGlnHisGly	780
Db	2647	GGCTACGACACAGGGTCACTGAGCGGTACATGAGAGTCTCTGAGAACACACAGCACGGC	2706
QY	781	TyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeu	800
Db	2707	TATGAGCGGGGTTCCGTGGCCCTGACAGTGGAGAAACTGCCCAATATGAGCCCAACCGCTTG	2766
QY	801	LeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVal	820

```
|||||
Db 2767 CTTATCCCGCAGGGCTTCCTCGAGCAAAACGAGCACTTTTCCACAAACCTTCCTCGTC 2826
Qy 821 SerIleuIleAtrgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHis 840
Db 2827 TCCCAAGATCCGAGCGAGGAAACCTTACCAAGCTCCAGATCTACCCCAAGAGAGACAC 2886
Qy 841 SerIleArgCysProGluSerGlyGlnHisTyrGluValThrLeuLeuHisPheLeuGln 860
Db 2887 AGATATCCCTGCCCCGAGCTGGGCGAGCAGTATGAATGACAGTTGCTGCTGCTACAG 2946
Qy 861 GluTyrIleu 863
Db 2947 GAATACCTC 2955

RESULT 4
US-09-976-674-36
; Sequence 36, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Q1, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976, 674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240, 117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 4180
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-36

Alignment Scores:
Pred. No.: 0 Length: 4180
Score: 4558.50 Matches: 850
Percent Similarity: 98.49% Conservative: 0
Best Local Similarity: 98.49% Mismatches: 0
Query Match: 98.12% Indels: 13
DB: Gaps: 1

US-09-976-674-3 (1-863) x US-09-976-674-36 (1-4180)
Qy 1 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspAspPro 20
Db 367 ATGGCCACACCGGAGCCCAACGCGCAGCGAGCGAGCGCCGACAGATGACCGC 426
Qy 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
Db 427 GCCCGCCCTTCCAGAGTGCACAAAGCAGCTGTGGAGCGGCTCCGAGAGATATCCAGCGC 486
Qy 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 487 AGCGCAAGTACTCGGGGCTCATTTGTCAACAAGCGGCCCAAGCACTTCAGATTGTGCAG 546
Qy 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
Db 547 AAGAGGATGTGGTGGGCCCACTCCACCGCCTTACTACTCGGGAATGCCATATGGC 606
Qy 81 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
Db 607 AGCCGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAAAGGTTCCGAAAGAGGCTTG 666
Qy 101 LeuLeuLeuSerTyrLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
Db 667 CTGCTCTGCTCTGGAAGACATGCTGATCATTTCCAGGCGACGCCCAACATGGGCTC 726
Qy 121 TyrSerArgGluGluGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr 140
```

```
|||||
Db 727 TACTCTCCGGAGGAGGAGCACTCTGAGCGAGGAAACCCCTGGGGGCTTCGGCATCAC 786
Qy 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
Db 787 TCTTAGAGCTTCCACAGGAGAGTGGCTCTTCTCTTCCAGGCCAGCAAGACCTCTTC 846
Qy 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
Db 847 CACTGCCCGCAGCGGGGCAAGAGCGCTTCATGGTGTCCCTATGAAACCCCTGGAATTC 906
Qy 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
Db 907 AAGACCCAGTCTCAGGGCCCCCGGATGAGCCCAAAATCTGCCCTGCCAGCTCCCTTC 966
Qy 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg 220
Db 967 TTCTCTTCATCATTAACAGAGCACTGTGGGTGGCAATCGACATCGACGAGGAGCGG 1026
Qy 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240
Db 1027 CGGCTGACCTTCTGCAACCAAGTTTATCCATATGCTCGATGACCCCAAGTCTGGGGT 1086
Qy 241 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpCysPro 260
Db 1087 GTGGCACCTTCGTATACAGGAAGATTCGACCGCTTCACTGGGTACTGGTGGTCCCC 1146
Qy 261 ThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluVal 280
Db 1147 ACAGCTCTCTGGAAAGTTTCAGAGGCGCTCAAGACGCTGCAAACTCTGTAGAGAACTC 1206
Qy 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArgLysThr 300
Db 1207 GATGAGTCCGAGGTGGAGTATTCACGCTCCCTCTCTGGGCTAGAAAGAAAGACG 1266
Qy 301 AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320
Db 1267 GACTCGTATCGGTACCCCGAGCAGGAGCAAGAAATCCCAAGATGCTCTGAAACTGGCT 1326
Qy 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340
Db 1327 GAGTTCCAGACTGCAGCCAGGCGCAAGATCTCTCGACCCAGAGAAAGACCTGGTCAG 1386
Qy 341 ProPheSerSerLeuPheProLysValGluThrIleAlaArgAlaGlyTrpThrArgAsp 360
Db 1387 CCGTTACAGCTCGTGTGCCAAGGTGGATACATCGCCAGGCGGGGTGACCCGGGAT 1446
Qy 361 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 380
Db 1447 GGCAAATACGCGCTGGGCTATGTTCTGTGACCGGCCCGCAGCTGCTCAGCTCTCTC 1506
Qy 381 LeuProProAlaLeuPheIleProSerThrGluAsnGluGluArgLeuAlaSerAla 400
Db 1507 CTCCCCCGGCGCTGTTTCATCCGACGACAGAGATGAGAGACGCGCTGCTCTGCC 1566
Qy 401 ArgAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTrp 420
Db 1567 AGAGCTGTGCCAGAAATGTCCAGCCGTAATGTGTGAGGAGGATCCACCAAGCTGTGG 1626
Qy 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGlyLysGluAspGluLeu 440
Db 1627 ATCATATGTTCAATGACATCTTATCCCTTCCCAATTCAGAGGAGAGACGACTCTGCG 1686
Qy 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460
Db 1687 TTTCTCCGCGCAATGAAATGAAAGCCGCTTCTGCTTCTGTAACAAAGTACCGCGCTT 1746
Qy 461 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGlyGluAspGluPheLys 480
Db 1747 TTAATAATCCAGGGCTAGATGGAGTGAGCCCTTCAGCCCGGAGAAAGATGAAATTAAG 1806
Qy 481 CysProIleLysGluGluIleAlaLeuThrSerGlyLysTrpGluValLeuAlaArgHis 500
```

Db 1807 TGCCCATTAAGAGAGATTGCTCTGACACGCGTGAATGGAGTTTGGCGAGCAC 1866
Qy 501 GlySerLysIleTrpValAsnGluGlnThrLysLeuValTyrPheGlnGlyThrLysAsp 520
Db 1867 GGCTCCAG-----GGCACCAAGCAC 1887
Qy 521 ThrProLeuGlnHisIleLeuTyrValSerTyrGluAlaIleGluIleValArg 540
Db 1888 ACGCGGTGAGACACACCTGACGTGACATGAGAGCGCGGAGATGCTACGC 1947
Qy 541 LeuThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
Db 1948 CTCACACAGCGCGGCTTCCATGCTGCTCCATGAGCAGCAAACTGACATGTTCCGC 2007
Qy 561 SerHisTyrSerSerValSerThrProProGlyValHisValTyrLysLeuSerGlyPro 580
Db 2008 AGCCACTACAGACGCTGAGCAGCGCGCGCTGCGACGCTGACAGCTGAGCGGCGCC 2067
Qy 581 AspaSarpProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaIleAsp 600
Db 2068 GACGACGACCGCTGACACAGACGCGCTTCTGCGCATGATGAGGACAGCAC 2127
Qy 601 CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 620
Db 2128 TGCCCCCGGATTATGTTCTCTCCAGAGATCTTCATTCCACACGCGCTCGGATGCGCG 2187
Qy 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal 640
Db 2188 CTCATGGGATGATCTACAGACCGCCAGCGCTGACAGCAGGAGAAAGACCCACCGTC 2247
Qy 641 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
Db 2248 CTCCTTGTATATGAGAGCGCGCGCGCTGCGATGAAATCTCTTCAAGCATCAAG 2307
Qy 661 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArg 680
Db 2308 TACTTGGCGGCTCACACACTGCGCTCCGCGGCTAGCGCGCTGTTGATGACGAGCAC 2367
Qy 681 GlySerCysGlnArgGlyLeuArgPheGluGlnAlaLeuLysAsnGlnMetGlyIleVal 700
Db 2368 GGCTCTCTACAGCGAGGCTTGGTTCGAAAGGCGCGCTGAAACCAATGGGCGACGG 2427
Qy 701 GlnIleGluAspGlnValGlnGluLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
Db 2428 GAGATCGAGAGACAGCTGAGAGGCGCTGCACTGCGCGAGAAATGATGCTTCATCGAC 2487
Qy 721 LeuSerArgValAlaIleHisGlyTyrPheSerTyrGlyGlyPheLeuSerLeuMetGlyLeu 740
Db 2488 CTGACCGGAGTTGCCATCCATGGCTGCTGCTGAGGCGCTTCTCGCTCATGGGCGTA 2547
Qy 741 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMet 760
Db 2548 ATCCCAAGCGCGAGGTGTTCAAGGTGGCCATCGCGGCGCGCGCTGAGAG 2607
Qy 761 AlaTyrAspThrGlyTyrThrGlnArgTyrMetAspValProGlnAsnAsnGlnHisGly 780
Db 2608 GCGTACGACACAGGAGTACACTGAGCGCTACATGAGAGCTCCCTGAGAACACAGCACGCG 2667
Qy 781 TyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeu 800
Db 2668 TATGAGGCGGCTTCCGTGCGCTGACAGTGGAGAGCTCCCAATGAGCCCAACGCGTTG 2727
Qy 801 LeuIleLeuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuVal 820
Db 2728 CTTATCTTCACAGGCTCTCTGAGCAAAACGTCATTCTTCCACAAATCTCTCGTC 2787
Qy 821 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnArgHis 840
Db 2788 TCCCAAGTATCGGAGAGGAGAAACCTTACAGCTCCAGATCTACCCCAACAGAGACAC 2847
Qy 841 SerIleArgCysProGluSerGlyGluHisTyrGluValThrLeuLeuHisPheLeuGln 860
Db 2848 AGTATTCGCTGCCGAGCGGCGAGCACTATGAAGTACAGCTGCTGACATTTCTACAG 2907

Qy 861 GluTyrLeu 863
Db 2908 GAATACCTC 2916
RESULT 5
US-09-976-674-34
; Sequence 34, Application US/09976674
; Patent No. US20020115843A1
GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
; FIDE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976, 674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 4263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-34
Alignment Scores:
Pred. No.: 0 Length: 4263
Score: 4558.50 Matches: 850
Percent Similarity: 98.49% Conservative: 0
Best Local Similarity: 98.49% Mismatches: 0
Query Match: 98.12% Indels: 13
Db: 10 Gaps: 1
US-09-976-674-3 (1-863) x US-09-976-674-34 (1-4263)
Qy 1 MetalIaThrThrGlyTyrProThrAlaAspArgGlyAspAlaAlaThrAspAspPro 20
Db 367 ATGGCCACACACCGGAGACCCCAACGCGCGCGACGCGAGCGCGCACAGATGACCGG 426
Qy 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
Db 427 GCCGCCCGCTTCAGGTGCGAGAAACAATCTGTGGAGCGGCTCGGAGCATTCACAGCG 486
Qy 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 487 ACCCGAAGTACTCGGGCTCATTTGTCAACAAGGCGCGCCACAGACTCCAGTTGTGCGAG 546
Qy 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
Db 547 AAGAGGAGTACTGCGGCGCCACCTCCCTACTACTGAGATGCGGAAATGCGCATATGCG 606
Qy 81 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
Db 607 ACCCGAGAACTCCCTCTTACTCTGAGATTCACAAAGAGTCCGGAAGAAGGCTCG 666
Qy 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
Db 667 CTGCTCTCTCTCTGAGAGAGATGCTGATCTTCATTCACAGCGACGCCACCATGGGCTG 726
Qy 121 TyrSerArgGluGlnGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyTyrThr 140
Db 727 TACTCTCGGAGAGAGAGTCTGAGGAGCGGAAACGCTGCGGCTCTTCGCGATCAC 786
Qy 141 SerTyrAspPheHisSerCysLeuSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
Db 787 TCTTACGACTTCACAGCAGAGAGTGGCTTCTCTTCTTCCAGGACAGACAGCTCTTC 846
Qy 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
Db 847 CACTGCGCGGAGCGGCGGCAAGAGCGCTTCATGATGATGCTCCATGAACCGCTGGAATC 906

QY 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
 DB 907 AAGACCAGTCTCAGGGCCCGGATGAGACCCCAAAATCTCCCTCGGACCCTCCCTTC 966
 QY 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluTrpGlyGluGlnArg 220
 DB 967 TTCTCTTCATCAATTAACAGCGACCTGTGGTGGCCAACTGAGACAGCGGAGAGCGGG 1026
 QY 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValIleuAspAspProLysSerAlaGly 240
 DB 1027 CGGGTACCTTCTGCAACAGGTTTATCCATGCTCGGATGACCCCAAGCTTCCGGGT 1086
 QY 241 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTrpTrpCysPro 260
 DB 1087 GTGGCCACCTTCGTCAATACAGAGAGAGTTCAGCCGCTTCACTGGTACTGGTGGTGGCCC 1146
 QY 261 ThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTrpGluGlnVal 280
 DB 1147 ACACCTCTCTGGGAAGGTTTCAGAGGGCTCAAGAGCGTGCATTCCTGATCAGAGAGTTC 1206
 QY 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArgLysThr 300
 DB 1207 GATGATCCGAGGTGAGAGTCAATTCAGTCCCTCTCTCGCTAGAGAAAGAGAGCG 1266
 QY 301 AspSerTrpArgTrpProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320
 DB 1267 GACTCGTATCGGTACCCAGAGACAGGACAAAGATCCCAAGATTCCCTTGAACCTGGCT 1326
 QY 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340
 DB 1327 GAGTTCACGACTGCAGCAGGAGGCAAGATCTCTCGACCCAGAGAGAGACTGCTGGCAG 1386
 QY 341 ProPheSerSerLeuPheProLysValGluTrpIleAlaArgAlaGlyTrpThrArgAsp 360
 DB 1387 CCTTCACGCTGCTGTTCCCAAGGTGAGTACATCGCCAGGGCCGGGTGAGACCCGGGAT 1446
 QY 361 GlyLysTrpAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 380
 DB 1447 GCGAAATACGCGCTGGGCGCATGTCCTGGACCGGGCCAGAGTGGCTCAGCTGCTC 1506
 QY 381 LeuProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAla 400
 DB 1507 CTCCCCCGGCGCTGTTCATCCGAGCACAGAGATGAGGAGCGGCTGACCTCTGCC 1566
 QY 401 ArgAlaValProArgAsnValGlnProTrpValValTrpGluGluValThrAsnValTrp 420
 DB 1567 AGACCTGCCCCAGGAATGTCCAGCCGTATGTGTGTACGAGAGGTACCAACGCTCTGG 1626
 QY 421 IleAsnValHisAspIlePheTrpProPheProGlnSerGluGluLysAspGluLeuGly 440
 DB 1627 ATCAATGTTCATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGAGAGCTCTGC 1686
 QY 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTrpValThrAlaVal 460
 DB 1687 TTCTCTCGCGCCATGATGACAGACCGCTTCTGCCATTTGTACAAAGTACCGCCCTT 1746
 QY 461 LeuLysSerGlnGlyTrpAspTrpSerGluProPheSerProGluGluLysAspGluPheLys 480
 DB 1747 TTTAAATCCCGAGGCTACGATGAGTGAAGCCCTTCACGCCCGGAGAGATGAAATTTAG 1806
 QY 481 CysProIleLysGluGluIleAlaLeuThrSerGluTrpGluValLeuAlaArgHis 500
 DB 1807 TGCCCATTTAGGAAGATGTCTGACCAAGCGGTGAAATGGAGGTTTGGCGAGGAGAC 1866
 QY 501 GlySerLysIleTrpValAsnGluGluThrLysLeuValTrpPheGlnGlyThrLysAsp 520
 DB 1867 GGCCTCCAG-----GGCACCACCAAGAGC 1887
 QY 521 ThrProLeuGluHisLeuTrpValValSerTrpGluAlaGluGlnIleValArg 540
 DB 1888 ACGCGCGTGAGCACCACTCTACGTGGTACGATATGAGGCGCGGAGATGCTACGCG 1947

QY 541 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
 DB 1948 CTCACACAGCCCGGCTTCTCCATAGTGTCTCCATGAGACCAGAACTTGCATGATGTGCTC 2007
 QY 561 SerHisTrpSerSerValSerThrProCysValHisValTrpLysLeuSerGlyPro 580
 DB 2008 AGCCACTACACACAGCTGAGCAGCCCGCTGCTGACGCTCAAGCTCAGCAGCGGCCCC 2067
 QY 581 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaSer 600
 DB 2068 GACGACGACCCCTGACACAGACAGCCCGCTTCTGCGCTACCATGATGAGAGCAGCCAGC 2127
 QY 601 CysProProAspTrpValProProGluIlePheHisPheHisThrArgSerAspValArg 620
 DB 2128 TGCCCCCGGATTAATGTTCTCCAGAGATCTTCCATTTCACAGCGCGCTGGATGTGGG 2187
 QY 621 LeuTrpGlyMetIleTrpLysProHisAlaLeuGlnProGlyLysLysHisProThrVal 640
 DB 2188 CTTCACGCGCATGATCTACACAGCCCGCTGCTGACGCGAGGAGAGAGCACCACCGCTC 2247
 QY 641 LeuPheValTrpGlyGlyProGluValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
 DB 2248 CTCTTTGTATATGAGAGGCCCCCAGGTGAGTGGTGAATTAATCTCTTCAAGGCATCAG 2307
 QY 661 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTrpAlaValValIleAspGlyArg 680
 DB 2308 TACTTGGCGCTCAACACACTGCTCCCTCGGTACGCGGTGATGATGAGCGGAGG 2367
 QY 681 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnVal 700
 DB 2368 GGCCTCTGTACAGAGGCGCTTGTGCAAGGGGCCCTGAAATATGGCCAGGTG 2427
 QY 701 GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTrpGlyPheIleAsp 720
 DB 2428 GAGATCAGAGACCAAGTGAAGGCTCGCATGCTGCTGGCCGAGAAATGATGTATGACAC 2487
 QY 721 LeuSerArgValAlaIleHisGlyTrpSerTrpGlyGlyPheLeuSerLeuMetCylLeu 740
 DB 2488 CTGAGCGGAGTGGCATTCATGAGTGGTGTCTACGAGGGGCTCTCTGCTCATGGGCTA 2547
 QY 741 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMet 760
 DB 2548 ATCCACAAGCCCCAGGTGTTCAAGGTGACATGCGCGGGTCCCGGTACCCGTGTGATG 2607
 QY 761 AlaTrpAspThrGlyTrpThrGluArgTrpMetAspValProGluAsnAsnGlnHisGly 780
 DB 2608 GCTTACGACACAGGCTACAGTACGACGCTACATGACGCTCTGAGAACACACACACGCGC 2667
 QY 781 TyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeu 800
 DB 2668 TATGAGGCGGGTTCGTTGCGCTCACGTGAGAGAGTGGCCCAATGAGCCCAACCGCTTG 2727
 QY 801 LeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAspPheLeuVal 820
 DB 2728 CTATATCTCCAGCGCTTCTGAGGAAACGTGACATTTTCCACACAACTTCCCTGCTC 2787
 QY 821 SerGlnLeuIleArgAlaGlyLysProTrpGlnLeuGlnIleTrpProAsnGluArgHis 840
 DB 2788 TCCCACTGATGCCAGCAGCAGGAAACCTTACAGCTCCAGATCTCAACCCAGAGAGACAC 2847
 QY 841 SerIleArgCysProGluSerGlyLysHisTrpGluValThrLeuLeuHisPheLeuGln 860
 DB 2848 AGTATTCGCTGCCCGGAGTGGGAGACACATATGAAATGACAGTGTGCACTTTCTACAG 2907
 QY 861 GluTrpLeu 863
 DB 2908 GAATACCTC 2916

RESULT 6
 US-09-976-674-32
 ; Sequence 32, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:


```
Db 2107 GAGGACGACCCCTGACAAAGACAGCCCGCTTCTGGCGCTAGCATGATGAGGACGACCCAGC 2166
Qy 601 CysProbaAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 620
Db 2167 TGGCCCCGGGATGTTATGTTCCCGCAGAGATGTTCCATTTCACACAGCCCTGGAGTGTGGG 2226
Qy 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysHisProThrVal 640
Db 2227 CTCACGCGCATGATCTCAAGCCCGCCCTTGCAGCCAGGAGGAAGACACCCACCGCTC 2286
Qy 641 LeuPheValTyrGlyLysProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
Db 2287 CTCCTTGTATATGAGGAGCCCGCAGAGTCAAGTGTGATTAATCTCTTCAAGGCAATCAG 2346
Qy 661 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValAlaIleAspGlyArg 680
Db 2247 TACTTGGGGGTCAACACACTGCGCTCCCTGGCGTACCGCGTGTGATGATGAGGCGAGC 2406
Qy 681 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnVal 700
Db 2407 GGCCTCTGTAGAGGAGGCTTGGCTTCGAAGGGGCGCTGAAACCAAAATGGGCGAGGTG 2466
Qy 701 GluIleGlnAspGlnValGlnGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
Db 2467 GAGATCGAGGACAGGTGGAGGGCTGCAGTTCGTGGCCGGAAGTATGGCTTATCGAC 2526
Qy 721 LeuSerArgValAlaIleHisGlyTyrSerTyrGlyGlyPheLeuSerLeuMetGlyLeu 740
Db 2527 CTGAGCGGAGTGGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2586
Qy 741 IleHisLysProGlnValPhe-LysValAlaIleAlaGlyAlaProValThrValTrpMe 760
Db 2587 ATCCACAGCCCGCCGCTTTCAGAG----- 2611
Qy 760 TalatYrAspThrGlyTyrThrGlnArgTyrMetAspValProGlnAsnGlnHisGly 780
Db 2611 ----- 2611
Qy 780 YTYrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGlnProAsnArgLe 800
Db 2612 -----CCCAACCGCTT 2622
Qy 800 ULeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVa 820
Db 2623 GCTTATCTCCACAGGCTTCTGAGCAAAAGCTGCTTTTTCACACAACTTCCCTGCT 2682
Qy 820 LserGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnArgHis 840
Db 2683 CTCACACTGATCGGACGAGGAGAACTTACAGCTCAGATCTAACCCCAACGAGAGACA 2742
Qy 840 sSerIleArgCysProGlnSerGlyGlnHisTyrGlnValThrLeuLeuHisPheLeuG 860
Db 2743 CAGATATCGCTGGCCCGAGTCGGCGAGCAGCACTATGAGTCACTGCTGCACTTCTACA 2802
Qy 860 ngJuTyrLeu 863
Db 2803 GGAATACCTC 2812

RESULT 7
US-09-976-674-30
; Sequence 30. Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
```

```
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 4159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-30

Alignment Scores:
Pred. No.: 0 Length: 4159
Score: 4337.00 Matches: 814
Percent Similarity: 94.33% Conservative: 1
Best Local Similarity: 94.21% Mismatches: 0
Query Match: 93.35% Indels: 49
DB: Gaps: 1

US-09-976-674-3 (1-863) x US-09-976-674-30 (1-4159)

Qy 1 MetalatThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspAspPro 20
Db 367 ATGGCCACACCGGAGACCCCAACGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 426
Qy 21 AlalaArgPheGlnValGlnLysHisSerThrAspGlyLeuArgSerIleIleHisGly 40
Db 427 GCGCGCGCTTCCAGGTGCAAGAACACTCGTGGGACGGGCTCGGACCATCATCCACGGC 486
Qy 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 487 AGCCGCAAGTACTGGGCTCATTTGTACACAGCGCCGCCACAGCTTCCAGTTTGTGCAG 546
Qy 61 LysThrAspGlnSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
Db 547 AAGACGATAGTCTGGGCGCCCACTCCACCGCTTACTACTGCGGAATGCGATATGCGC 606
Qy 81 SerArgLysAsnSerLeuLeuTyrSerGlnIleProLysLysValArgLysGluAlaLeu 100
Db 607 AGCCGAGAGAACTCCCTCTTACTGTGAGATTCACAAAGGTCGGAAGAGGCTCTG 666
Qy 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
Db 667 CTGCTCCTGCTCTGGAAGCAGATGCTGATCATTTCCAGGCGCAAGCCCAACATGGGGTCT 726
Qy 121 TyrSerArgGluGluGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr 140
Db 727 TACTCTGGGAGAGGAGCTGCTGAGGAGCGGAAACGCTGGGGGCTTCCGGCATCAC 786
Qy 141 SerTyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaAsnSerLeuPhe 160
Db 787 TCTTACAGACTTCCACAGCGAGAGTGGCTTCTCTTCCAGGCGCAGACAGAGCTTCTC 846
Qy 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
Db 847 CACTGCGCGCAGCGCGCAAGAACGCTTCATGGTGTGCCCTATGAACCGCTGGAATC 906
Qy 847 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
Db 907 AAGACCCAGTGTCTGAGGCGCCCGGATGAGCCCAAAATCTGCCCTGGCGCAGCTTCC 966
Qy 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluArg 220
Db 967 TTCTCTTCTATCAATTAACAGGACTGTGGTGGCCAACTCCAGACAGCGAGAGCGG 1026
Qy 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240
Db 1027 CGGTGACCTTTCGCCCAAGGTTTATCCAAATGCTCTGATGATGCCCAAGTGTGGGGT 1086
Qy 241 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpCysPro 260
Db 1087 GTGGCCACTTTCGATACAGGAAGATGAGCGGCTTCACTGGGTACTGGTGTGCCCC 1146
Qy 261 ThrAlaSerTrpGluLysArgLysGlyLeuLysThrLeuArgIleLeuTyrGluGluVal 280
Db 1147 ACAGCTCTCTGGGAAGGTTCAGAGGGCTTCAGACGCTGCAATCTGTATGAGGAATC 1206
```


QY 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArgLysThr 300
 Db 1207 GATGAGTCGAGAGTGAGAGTACGTACCTCCCTCCGCTGAGAAAGAGAGACG 1266
 QY 301 AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320
 Db 1267 GACTCGTATCGGTACCCAGGACAGGACAAAGATCCCAAGATGCTTGAACTGGCT 1326
 QY 321 GluPheGluThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340
 Db 1327 GAGTTCACAGCTGACAGGACGAGGCAAGATCGCTGACCCAGCAGCAAGAGAGCTGTGAG 1386
 QY 341 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyThrProArgAsp 360
 Db 1387 CCCTTCACACTGCTCTTCCCGAGGTGAGATACATCCCGAGGCGCGGTGGACCCGGGAT 1446
 QY 361 GlyLysTyrAlaTyrAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 380
 Db 1447 GGCAAAATACGCGCTGGCCCATGTTCTGGACCGCGCCAGCAGAGTGCTCCAGCTCGCTGC 1506
 QY 381 LeuProProAlaLeuPheIleProSerThrGluAsnGluGluArgLeuAlaSerAla 400
 Db 1507 CTCCCGCGCGCTGCTTCATCCCGAGCAGAGAAATGAGGAGCAGCGCTACCTCTGCG 1566
 QY 401 ArgAlaValProArgAsnValGlnProTyrValIleTyrGluGluValThrAsnValTrp 420
 Db 1567 AGAGCTGTCCCGAGAGATGCCGATGTGGTGTACGAGGAGGTCCACACGCTGTGG 1626
 QY 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluLysAspLysLeuLys 440
 Db 1627 ATCAATGTTATGACATCTTATCCCTCCCGCAATCAGAGAGAGAGAGAGAGCTGCG 1686
 QY 441 PheLeuAlaGluAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460
 Db 1687 TTTCTCCCGCGCAATGCAATGCAAGACCGGCTCTCCATTTGTACAAAGTCAACCCCGTT 1746
 QY 461 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGluGluAspLysPheLys 480
 Db 1747 TTAATATCCCGAGGCTGACATTTGAGTGTAGCCCTTCAGCCCGGGGAGAGATATAATTTAAG 1806
 QY 481 CysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpGluValIleLeuAlaArgHis 500
 Db 1807 TGCCCATTAAGAGAGAGAGTTGCTGTACACAGCGGTGATGGAGGTTTGGCGAGGCAC 1866
 QY 501 GlySerLysIleTyrValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp 520
 Db 1867 GGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCACAGGCGCACCAAGGAC 1926
 QY 521 ThrProLeuGluHisHisLeuTyrValIleSerTyrGluAlaAlaGlyIleValArg 540
 Db 1927 AGCGCGCTGGAGCACACCTTACGTGTGACCTTGAAGCGCGCGGCGAGATCTGTACC 1986
 QY 541 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
 Db 1987 CTCACACAGCGCGGCTTCTCCATAGCTGTCCATGAGCAGACATTCACATGTGTCGTC 2046
 QY 561 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 580
 Db 2047 AGCCACTACAGCAGGTGAGCAGCGCGCGCTGTCAGCTGTACAAAGCTGAGCGCGCCC 2106
 QY 581 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetCysIleAlaAsp 600
 Db 2107 GACGAGCAGACCCCTGCACACAGACGCCCTTGTGGCTAGCATGTGAGGACGACGACGAC 2166
 QY 601 CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 620
 Db 2167 TGCCCCCGGATTAATGTCTCCAGAGATCTTCATTTCCACACGGGCTCGGATGTGGCG 2226
 QY 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal 640
 Db 2227 CTCTACGGCATGATCTACAAAGCCACAGGCTTGACAGCCGAGGAGAAAGACCCACCGCTC 2286

QY 641 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
 Db 2287 CTCTTTGATATGAGAGGCCCCAGGTGCAGCTGGTGAATTACTCTTCAAGAGCATCAG 2346
 QY 661 TyrLeuAlaGluAsnThrLeuAlaSerLeuGlyTyrAlaValAlaValIleAspGlyArg 680
 Db 2347 TACTTGGCGCTCAACACACTGCTGCCCTCGGCTACGCGCTGTGTGATGTGACGCGCAG 2406
 QY 681 GlySerCysGlnArgGlyLeuArgPheGluGluGlyAlaLeuLysAsnGluMetGlyValVal 700
 Db 2407 GGCTCTGTCAAGGAGGCTTGTGTTCCAAAGGGGCCCTGAAAACCAATGGGCCAGGTG 2466
 QY 701 GluIleGluAspGluValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
 Db 2467 GAGATCGAGAGACAGGTGAGGGGCTCGCAGTTCTGTGGCCGAGAAATATGGCTTCATCAG 2526
 QY 721 LeuSerArgValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeu 740
 Db 2527 CTGAGCCGAGTGGCCATCATGCTGTGTCTTACGGGGGCTTCTCGCTCATGGGGCTA 2586
 QY 741 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMe 760
 Db 2587 ATCCACAAGCCCCAGGTGTCAAG----- 2611
 QY 760 ValTyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGly 780
 Db 2611 ----- 2611
 QY 780 YTrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLe 800
 Db 2612 -----CCCAACCGGTT 2622
 QY 800 uLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAspPheLeuVal 820
 Db 2623 GCTTATCTCCACGCGCTTCGTGGAGAAACCTGTGCACTTTTCCACAAACTTCCTGCT 2682
 QY 820 LserGlnLeuIleArgAlaGlyLysProTyrGlnGlnGlnIleTyrProAsnGluArgHis 840
 Db 2683 CTCCCACTGATCCGAGCAGGAGAAACCTTACAGCTCCAGATCTACCCCAAGAGAGACA 2742
 QY 840 sSerIleArgCysProGlnSerGlyLysIleHisTyrGluValThrLeuHisHisPheLeuGly 860
 Db 2743 CAGTATTCGCTGCCCGCGAGATCGGCGAGCAGCATATGAAATGACAGCTTGTCTTACA 2802
 QY 860 nGluTyrLeu 863
 Db 2803 GGAATACCTC 2812

RESULT 8
 US-09-976-674-40
 ; Sequence 40, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akinsanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Julien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DDPV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; PRIOR APPLICATION NUMBER: 2001-10-12
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 40
 ; LENGTH: 4037
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-976-674-40

Alignment Scores: 0 Length: 4037
 Pred. No.:

Score: 4249.50 Matches: 801
Percent Similarity: 92.82% Conservative: 1
Best Local Similarity: 92.71% Mismatches: 0
Query Match: 91.47% Indels: 62
DB: 10 Gaps: 2

US-09-976-674-3 (1-863) x US-09-976-674-40 (1-4037)

QY 1 MetaIaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaIaThrAspPro 20
|||||
Db 367 ATGGCCACACACGGGAGCCCAAGCCGAGCCGAGCGACGCGCCACAGATGACCGG 426

QY 21 AlaIaIaArgPheGlnValGlnLysHisSerTyrAspGlyLeuArgSerTyrIleHisGly 40
|||||
Db 427 GCCCGCCGCTTCAGAGTGCAGAACACCTCGTGGAGCGGGCCGAGCATCATCCACGCG 486

QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
|||||
Db 487 AGCCGCAGATCTCGGGCTTATGTCAACAGGCGGCCCGCAGACTTCACATTTGTGAG 546

QY 61 LysThrAspGlySerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
|||||
Db 547 AAGACGATGATGTCGGGCCCCACTCCACCGCCTTACTACTGGAATGCCATATGCGC 606

QY 81 SerArgGlyAsnSerLeuLeuTyrSerGluIleProLysValArgLysGluAlaLeu 100
|||||
Db 607 AGCCGAGAGAACTCCCTCTACTGTGAGATCCCAAGAGGTCCGGAAGAGGCTCTG 666

QY 101 LeuLeuLeuSerTyrPlyGlnMetLeuAspHisPheGlnAlaThrProHisGlyVal 120
|||||
Db 667 CTGCTCTCTCTGGAACAGATGCTGATCATTTCCAGGCCACGCCCCACCATGAGGCTC 726

QY 121 TyrSerArgGlyGluGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr 140
|||||
Db 727 TACTCTGGGAGAGAGCTCTGAGGAGCGGAACGCTGGGGGCTTCGGCATCACC 786

QY 141 SerTyrAspPheHisSerGlySerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
|||||
Db 787 TCTACAGACTTCCACAGCGAGAGTGGCTCTCTCTCCAGGCGAGCAAGACGCTCTTC 846

QY 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
|||||
Db 847 CACGCGCGGAGCGGCGGAGAACGGCTTCATGTGTCTCCATTGAAACCTCTGAAATC 906

QY 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
|||||
Db 907 AAGACCCAGTCTCAGGAGCCCGGATGAGACCCCAAAATCTGCCCTGCCGACCTTGCTTC 966

QY 201 PheSerPheIleAsnAsnSerAspLeuTyrValAlaAsnIleGluThrGlyGluArg 220
|||||
Db 967 TTTCTCTTCATCAATTAACAGGAGCTGTGGGTGGCAACATCGAGACAGGCGAGAGCGG 1026

QY 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240
|||||
Db 1027 CGGCTGACCTTCTCCACCAAGGTTTATCCATGTCTCTGGATGACCCCAAGCTTGGCGGT 1086

QY 241 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTyrProCysPro 260
|||||
Db 1087 GTGGCCACCTTCATACAGAGAGAGTTCAGCCGTTCACTGGGTGCTGGTGGCCCC 1146

QY 261 ThrIaIaSerTyrPglLysGlySerGlyLeuLysThrLeuArgIleLeuTyrGluVal 280
|||||
Db 1147 ACACCCCTCCGGAAGGTTTCAGAGGCTTCAGAGCGCTGCAAACTCTGTATGAGAGATC 1206

QY 281 AspIaSerGlyValGluValIleHisValProSerProAlaLeuGluArgLysThr 300
|||||
Db 1207 GATAGTCCGAGGTGGAGGTATTCACGCTCCCTCTGCGCTAGAGAAAGAGAAAGC 1266

QY 301 AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320
|||||
Db 1267 GACTCGATCGGTACCCAGAGACAGCAAGAAATCCCAAGATGGCTTGAATCGGCT 1326

QY 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340

|||||
Db 1327 GAGTTCAGACTGCACAGCCAGGCAAGATCGTCTCGACCCAGGAGAGAGACTGTGCAG 1386

QY 341 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrProThrAsp 360
|||||
Db 1387 CCTTCAGCTCGCTGTTCCTCCAGAGGTGAGTACATCGCAGGGCGGGGTGAGCCGGGAT 1446

QY 361 GlyLysTyrAlaTyrPalaMetPheLeuAspArgProGlnGlnTyrPleuGlnLeuValLeu 380
|||||
Db 1447 GGCAAAATAGCGCTGGGCGATGTCTCGAGCGGCCAGGAGTGCGGTCCACGCTGCTTC 1506

QY 381 LeuProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAla 400
|||||
Db 1507 CTCCCGCGCGCTTCATCCAGACAGAGATAGAGGCGCGCTAGCCTTGCC 1566

QY 401 ArgAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTyr 420
|||||
Db 1567 AGAGCTGTCCCGAGGAATGTCACACCGTATGTGGTGTACAGAGGTCCACCAACGCTTGC 1626

QY 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGlyGluAspGluLeuGly 440
|||||
Db 1627 ATCAATGTTCATGACATCTTCTATCCCTTCGCCCAATCAGAGGAGAGAGAGACTGTGC 1686

QY 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460
|||||
Db 1687 TTTCTCCGCGCAATGAATGACAGACCGGCTTGCATTTGTCAAAAGTCCACCGCGTT 1746

QY 461 LeuLysSerGlnGlyTyrAspTyrSerGluProPheSerProGlyGlyLysAspGluPheLys 480
|||||
Db 1747 TTTAAATTCGCCAGGCTACGATGTGAGTGAGCTTTCAGACCCCGGAGAAATGAATTTAAG 1806

QY 481 CysProIleLysGluGluIleAlaLeuThrSerGlyGluTyrPglValLeuAlaArgHis 500
|||||
Db 1807 TGCCCAATTAAAGAAAGATGTCTGACACGCGTGAATGGAGGTTTGGCGAGGCAC 1866

QY 501 GlySerLysIleTyrPalaAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp 520
|||||
Db 1867 GGCCTCCAG-----GGCACCAAGGAC 1887

QY 521 ThrProLeuGlnHisHisLeuTyrValValSerTyrGluAlaAlaGlyLysIleValArg 540
|||||
Db 1888 AGCGCGGTGAGACACACCTTACGTGTGACGATGAGCGGCGGCGAGATGTGACCC 1947

QY 541 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
|||||
Db 1948 CTCACACGCGCGGCTCTCCATAGCTGCTCATGAGCAGACTTCGACATGTTCGTC 2007

QY 561 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 580
|||||
Db 2008 AGCCACTACAGCAGCGTGAACGCGCGCTGCTGACGCTTACAGCTGAGCTGAGCGGCC 2067

QY 581 AspAspAspProLeuHisLysGlnProArgPheThrPalaSerMetGluAlaAlaSer 600
|||||
Db 2068 GAGCAGACACCCCTCGCAGCAAGACAGCCCGCTTGGGCTAGCATGATGAGAGCGACCA 2127

QY 601 CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 620
|||||
Db 2128 TGCCCCCGGATATGTTCCTCCAGAGATCTTCATTTCCACAGCGCGCTGGATGTGCGG 2187

QY 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal 640
|||||
Db 2188 CTCTACGCGATGATCTACAGGCCCGCTTGGAGCAGAGGAAGAACCCACCGCTC 2247

QY 641 LeuPheValTyrGlyLysProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
|||||
Db 2248 CTCTTTTATATGAGAGCGCCCGAGAGTGCACGTGTGAATATCTCTTCAAAAGCATCAAG 2307

QY 661 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArg 680
|||||
Db 2308 TACTTGGGCTCAACACACTGAGGCTCTCGTGGGCTTACCGCGGTGTGATGAGCGGAGG 2367

QY 681 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnVal 700
|||||

QY 381 LeuProAlaLeuPheIleProSerThrGluAsnGluGluAlaArgLeuAlaSerAla 400
 |||||
 Db 1507 CTCGCCCGCGCCGTGTCATCCCGACAGAGATGAGAGACAGCGCTGCTGCC 1566
 QY 401 ArgAlaValProArgAsnValGlnProIleValValTyrGluGluValThrAsnValTrp 420
 |||||
 Db 1567 AGAGCTGTCCCGAGAAATGTCACCGCTATGTGTGTACGAGAGAGGTACCAACCTGTGG 1626
 QY 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluGluAspGluLeuGly 440
 |||||
 Db 1627 ATCATATGTCATGACATCTTCTATCCCTCCCATACAGAGGAGAGAGAGCTCTGCC 1686
 QY 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460
 |||||
 Db 1687 TTTCTCCCGCCCAATGAAATGCAGACCGCTCTGCCATTGTACAAAGTCACCGCCGCT 1746
 QY 461 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGluGluAspGluPheLys 480
 |||||
 Db 1747 TTTAAATCCAGGGCTACGATGGATGGATGAGCCCTTCACGCCCGGGGAGATGAATTAAG 1806
 QY 481 CysProIleLysGluGluIleAlaLeuThrSerGlyLysTrpGluValLeuAlaArgHis 500
 |||||
 Db 1807 TGCCCATTAAGAGAGAGATGCTGTACACAGCGGTGAATGGAGTTTGGCGAGGCAC 1866
 QY 501 GlySerLysIleTrpValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp 520
 |||||
 Db 1867 GGCTCCAG-----GGCACCAAGAC 1887
 QY 521 ThrProLeuGlnHisIleSerTyrValValSerTyrGluAlaIleGluIleValArg 540
 |||||
 Db 1888 AGCGCGGTGGAGCACACCTTACGTGGTACGTATGAGGGCGCGGAGATGCTGTACG 1947
 QY 541 LeuThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
 |||||
 Db 1948 CTCACACAGCGCGCTTCTCCATAGCTGCTCCATGAGCCGAACTTGCATGTTCGTC 2007
 QY 561 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 580
 |||||
 Db 2008 AGCCACTACACACCGTACGACACCGCGCTGCGTGCACGCTCAAGCTGAGCGGCCCC 2067
 QY 581 AspAspAspProLeuHisIleGlnProArgPheTrpAlaSerMetMetGluAlaIleSer 600
 |||||
 Db 2068 GACACAGACCGCGTGCACACACCGCGCTGCGTGCACATGATGAGGACGACGAC 2127
 QY 601 CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 620
 |||||
 Db 2128 TGCCCCCGGATATGTTCCCTCCAGAGATCTTCATTTCCACAGCGCTCGGATGTGCGG 2187
 QY 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysHisProThrVal 640
 |||||
 Db 2188 CTCACGCGCATGATCTACAAACCCCGCTTCGACGACGAGGAGAAAGCACCCACGCTC 2247
 QY 641 LeuPheValTyrGluGlyProGluValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
 |||||
 Db 2248 CTCTTTGTATATGAGGCGCCCGAGGTGAGTGTGATTAATCTCTCAAGGCAATCAAG 2307
 QY 661 TyrLeuArgLeuAsnTrpLeuAlaSerLeuGlyTyrAlaValAlaIleAspGlyArg 680
 |||||
 Db 2308 TACTTGGGGGTCACACACTGTGCTCCCTGGGCTACCGCGGTGTGATTTAGAGGCGCAGG 2367
 QY 681 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnVal 700
 |||||
 Db 2368 GGCTCTGTGTAGCGAGGCTTCGTTGGAAGGGGCGCTGAATAAACCAATGGCGCAGGTG 2427
 QY 701 GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
 |||||
 Db 2428 GAGATCGAGGACAGGAGGAGGCTGTGAGTGTGCGCGGAGATATGGCTTATATGAC 2487
 QY 721 LeuSerArgValAlaIleHisGlyTyrSerTyrGlyLysPheLeuSerLeuMetGlyLeu 740
 |||||
 Db 2488 CTGAGCGGAGTGTGCGCATGAGTGTGCTCCTACGAGGGGCTTCTCTGCTATGTGGGCTA 2547

QY 741 IleHisLysProGlnValPhe-LysValAlaIleAlaGluAlaProValThrValTrpMe 760
 |||||
 Db 2548 ATCCACAAGCCCGAGGTGTCAAG----- 2572
 QY 760 TalatyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisG 780
 |||||
 Db 2572 ----- 2572
 QY 780 YTrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLe 800
 |||||
 Db 2573 -----CCCAACCGCTT 2583
 QY 800 uLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVal 820
 |||||
 Db 2584 GCTTATCTCCACGCGCTCTCGAGCAAAAGCTGACATTCTTCCACAAACTTCTCGT 2643
 QY 820 lSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluAlaG 840
 |||||
 Db 2644 CTCCCACTGATCCGAGCGAGGAAACCTTACACGCTCCAGATCTACCCCAACGAGAGCA 2703
 QY 840 sSerIleArgCysProGlnSerGlyGluHisTyrGluValThrLeuLeuHisPheLeuG 860
 |||||
 Db 2704 CAGATTTCGCTGCCCGGAGTGGCGGAGCAGCTATGAAATCAGTTTGTCTTCTACA 2763
 QY 860 nGluTyrLeu 863
 |||||
 Db 2764 GGAATACCTC 2773
 RESULT 10
 US-09-976-674-2
 : Sequence 2, Application US/0997674
 : Patent No. US20020115843A1
 : GENERAL INFORMATION:
 : APPLICANT: Qi, Steve
 : APPLICANT: Akinsanya, Karen
 : APPLICANT: Riviere, Pierre
 : APPLICANT: Junien, Jean-Louis
 : TITLE OR INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
 : FILE REFERENCE: 70669
 : CURRENT APPLICATION NUMBER: US/09/976, 674
 : PRIOR APPLICATION NUMBER: US 60/240, 117
 : PRIOR FILING DATE: 2000-10-12
 : NUMBER OF SEQ ID NOS: 61
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 2
 : LENGTH: 2671
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : US-09-976-674-2
 Alignment Scores:
 Pred. No.: 8, 28e-300
 Score: 2870.00
 Percent Similarity: 77.50%
 Best Local Similarity: 61.55%
 Query Match: 61.77%
 DB: 10
 Gaps: 2
 US-09-976-674-3 (1-863) x US-09-976-674-2 (1-2671)
 QY 24 PheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGlySerArgLys 43
 |||||
 Db 110 TTTTATGTTGAGCGGATTTCTCTGAGTCAAGCTTAATAAGCTGTCCGATACAGAAAA 169
 QY 44 TyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGlnLysThrAsp 63
 |||||
 Db 170 TATCATGCTACATAGAGGCTTACGACACATCATTTTCAATTTTGTGGAAGAGGATATG 229
 QY 64 GluSerGlyProHisSerHisArgLeuTyrTyrIleGluIleMetProTyrGlySerArgGlu 83
 |||||
 Db 230 CAGATGAGACTCTATGACACAGAAATTAATTACCTTCCATGTGTGAGAAACAGAAA 289

QY	544	ProIlyhSerHisSerCyssSerMetSerGlnAsnPhneAspMetPheValSerHisIytr	563
Db	1870	CGTGGTACACATATCTTGTCTGCACATCACTGACGACCTGGACCTTCTTTATGAAGAT	1929
OY	564	SerSerValSerThrProProCysValHisValIytrIyLysLeuSerGlyProAspAsp	583
Db	1930	AGTAAACAGAAAGATCCACACACTGTGTCTCCCTTTACAAAGCTTCAAGTCCGAAAGATGAC	1969
OY	584	ProLeuHisIySGlnProArgPheTrpAlaSerMetLeuAlaAlaSerCysProPro	603
Db	1990	CCAACTGTGCAAAACAAAGAAATTTTGGGCCACCATTTTGATGTACAGCAGTCCCTTCCCT	2049
OY	604	AspIytrValProProGlnIlePheHisPheHisIsthArgSerAspValArgLeuIyGly	623
Db	2050	GACATATCTCTCCAGAAATTTCTCTTTGAAAGTACTGATGATTTACATGTATGAGG	2109
OY	624	MetIleIytrIySPronHisAlaLeuGlnProGlyIyLysHisProThValLeuPheVal	643
Db	2110	ATGCTGTACAAAGCTCATATGATCTACAGCCTGGAAAGAAATATCCATACCTGTCTGTAT	2169
OY	644	TytrIyGlyIy--ProGlnValGlnLeuValAsnAsnSerPheIySGlyIleIyStryLeu	663
Db	2170	TATGTGTGTCTCCACAGGTCCAGTGTGTGAATAATCGTGTAAAGAGTCAAGATTTTCC	2229
OY	663	rgLeuAsnThrIleuAlaSerLeuGlyTytrValValValValIleAspIytrArgIySerc	683
Db	2230	GCTTGATATCCCTACGCTCTCTAGTATATGTGGTGTATGTATACAAACAGGGATCTT	2289
OY	683	IySGlnArgIyLeuArgPheGlnIyAlaIleuIyAsnGlnMetIyGlnValGlnIleG	703
Db	2290	GTCACCGAGGGCTTAATTTGAAGGGCGCTTTAAATATGAATGGGTCAATAGAAATG	2349
OY	703	IuAspGlnValGlnIyGlyLeuGlnPheValAlaGlnIyStryGlyPheIleAspIySera	723
Db	2350	ACGATCAAGGTGAAGGACCTCAATATCTAGCTTCCGATGTATTTCACTGACTTAGATC	2409
OY	723	rgValAlaIleHisIySGlyIyPyrSerTytrGlyIyPheLeuSerIleuMetGlyLeuHisI	743
Db	2410	GTCGGGCACTCCACGCGTGTCTTATGAGAGATACCTCTCCGTATGGCATTAATGACGA	2469
OY	743	IySProGlnValPheIySValAlaIleAlaGlyAlaProValThrValIyTrpMetAlaTyra	763
Db	2470	GGTCAGATATCTTCAGGGGTGCTATTTGCTGGGGCCCACTGCTGTGGATCTTATAG	2529
OY	763	SprHngIyTytrHngIyIyMetAspValProGlnAsnAsnGlnHisIyGlyTytrGln	783
Db	2530	ATACAGATACACGGAACGTTATATGGGTGCACCCCTGACACAGATACAAAGGGCTATTA	2589
OY	783	IagIySerValAlaIleHisIyValGlyIyLeuProAsnIyProAsnIyrgLeuIleuIel	803
Db	2590	TAGATCTGTGGCCATTCAGACGAAAGTATCCCTCTAACCAATGCTTTACTGCTCT	2649
OY	803	eunHisIyPheLeuAspGluAsnValHisPhePheHisIsthAsnPheLeuValSerGlnI	823
Db	2650	TACATGTCTTCTCGATGAGAAATGTCCATTTTGACATGACAGTATTTATACGATTTT	2709
OY	823	eunIleArgAlaGlyIyProTytrGlnLeuGlnIleIyTytrProAsnGlnArgHisSerIle	843
Db	2710	TAGGAGAGGCTGGAAACCCATATGATTTTACAGTCTATCTTACGAGAGACACAGCAAA	2769
OY	843	rgCysProGlnSerGlyGlnHisIyTytrGlyValThrIleuHisIsthPheLeuGlnIyTy	863
Db	2770	GAGTTCCTGAATCGGAGAACATTTATGAACTGCATCTTTTGCACATACCTCTCAAGAAAC	2829
OY	863	eu 863	
Db	2830	TT 2831	
RESULT 12			
US-09-976-674-22			
; Sequence 22, Application US/09976674			
; Patent No. US20020115843A1			

```

? GENERAL INFORMATION:
? APPLICANT: Qi, Steve
? APPLICANT: Akusanya, Karen
? APPLICANT: Riviere, Pierre
? APPLICANT: Junien, Jean-Louis
? TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
? FILE REFERENCE: 70669
? CURRENT APPLICATION NUMBER: US/09/976,674
? CURRENT FILING DATE: 2001-10-12
? PRIOR APPLICATION NUMBER: US 60/240,117
? PRIOR FILING DATE: 2000-10-12
? NUMBER OF SEQ. ID NOS: 61
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 22
?
? LENGTH: 4665
?
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-976-674-22

```

Alignment Scores:		
Pred. No.:	7.83e-275	Length: 4685
Score:	2642.50	Matches: 485
Percent Similarity:	73.10%	Conservative: 129
Best Local Similarity:	57.74%	Mismatches: 177
Query Match:	56.88%	Indels: 50
DB:	10	Gaps: 3

US-09-976-674-3 (1-863) x US-09-976-674-22 (1-4685)

OY		24	PheGlnValGlnIlyshSerrTraspolyleuAryserLlelleHslgSerArglys	43
Dd		316	TTTTATGTGGAGCGGATTCTCCGTGGACGTCACTTAAAAAGCGTCCTGCCGATACCAGAAA	375
OY		44	TyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGlnLysThrAsp	63
Dd		376	TATCATGGCTACATGATGGCTAAAGCACCATATGATTTTCATGCTTTTGGAAGAAGAT	435
OY		64	GluSerGlyProHisSerHisArgLeuTyrTyrLeuGlutMetProTyrGlySerArgGlu	83
Dd		436	CCAGATGGACCTCATTCACAGACAATCTATTACTCTGCACATGTCTGGAGAACAGAGAA	495
OY		84	AsnSerIleuLeuTyrSerGluIleProLysLysValArgLysGlnAlaIleLeuLeu	103
Dd		436	AATACACGCTTTTATTCTGAAATAATCCCAAACCTACATTAACAGACAGCTTTAATGCTC	555
OY		104	SerTrrLySGlnMetIleAspHisPheGlnAlaThrProHisLsgLysValTyrSerArg	123
Dd		556	TCTTGGAGCGCTTTTGATCTTTTTCAGGCAACACGTGACACTATGAAATGATTCTCGA	615
OY		124	GluGlnGluLeuLeuArgGluArgLysArgLeuGlnValPheGlyIleThrSerTyrAsp	143
Dd		616	GAAAGAAACATTAATTAAGAAAGAAAGAAAGCATTTGGACAGTCGGAATGCTTCTTAGCAT	675
OY		144	PheHisSerGluSerGlyLeuPheLeuPheGlnIleAserAsnSerLeuPheHisCysArg	163
Dd		676	TATCCCAAGCAAGAGGACACTTTCTGTTTAAAGCCGGTAAGTGAATTTATACAGTAAAA	735
OY		164	AspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIleLysThrGln	183
Dd		736	GATGAGGCGCCAACAAAGATTTATCCGACAAACCTTTAAAGGCCCATCTAAGTGAATAGT	795
OY		184	CysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPhe	203
Dd		796	TGTCCCAACATACGGATGATCTCAAAATTTAATGCCCTGTGATCCAGACTGGATTGGCTTTT	855
OY		204	IleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlnGluGlnArgLeuThr	223
Dd		856	ATACATACCAACAGATATTGGATATCTCAATCTCAATCGTAACACAGAAAGAGACCTCACT	915
OY		224	PheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGlyValAlaThr	243
Dd		916	TATGTGCAACATAGCATCTGCCAACATCGGAAGAAGATGCAATAGCTGGAGTGTGTACC	975

QY 244 PheValIleGlnGluPheAspArgPheThrGlyTyrTrpProCysProThrAlaSer 263
|||||.....
Db 976 TTGCTTCCAGAAAGATTTGATGATATCTGCTATTGGTGGTCCAAAAGCTGAA 1035
QY 264 TrpGlySerGlnGluLeuLysThrLeuArgIleLeuTyrGlnGluValAspGlnSer 283
|||||.....
Db 1036 ACAACTCCAGTGGTGGT---AAAATCTTAGAATTCATATGGAAGAAATGAGAAATCT 1092
QY 284 GluValGluValIleHisValProSerProAlaLeuGlnGluArgLysThrAspSerTyr 303
GAGGTGGAATATTATTCATGTTACATCCCTATGTTGGAAACAAGAGGCGCATTCATTC 1152
Db 1093 GAGGTGGAATATTATTCATGTTACATCCCTATGTTGGAAACAAGAGGCGCATTCATTC 1152
QY 304 ArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAlaGlnPheGln 323
1153 CGTATCTCTAAACAGGTACGCAATCTTAAGTCTTAAAGTCTGAGAAATGATG 1212
QY 324 ThrAspSerGlnGlyLysIleValSerThrGlnGluLysGlnLeuValGlnProPheSer 343
|||||.....
Db 1213 ATTGATCTGAGAGAGATCATAGATGTCATAGATTAAGCACTTAATTCACCTTTTGAG 1272
QY 344 SerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyr 363
|||||.....
Db 1273 ATTCTATTGGAAGAGATTGATATATTTGCCAGACCTGGATGACCTCGAAGGAAATAT 1332
QY 364 AlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProPro 383
1333 GCTTGTGCTACCTCTACTAGATCCGCTCCAGACTCGCTACAGATAGTGGATCTCACT 1392
QY 384 AlaLeuPheIleProSerThrGluAsnGlnGluGlnIleArgLeuAlaSerAlaArgAlaVal 403
|||||.....
Db 1393 GAATTAATTTATCCAGTAGAAGATGATGTTATGGAAAGGCAAGACATCTGATGACAGTG 1452
QY 404 ProArgAsnValGlnProTyrValValIleTyrGlnGluValThrAsnValTrpIleAsnVal 423
|||||.....
Db 1453 CCGTATCTGTCGAGCAGCACTAATATATCTATGAGAAGAAACAAGACATCTGATTAATATC 1512
QY 424 HisAspIlePheTyrProPheProGlnSerGlnGluLysGlnLeuCysPheLeuArg 443
1513 CATGACATCTTTCATGTTGTTCCCCAAAAGTCCAC---GAGAGGGAATTTGAGTTATTTT 1569
QY 444 AlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSer 463
|||||.....
Db 1570 GCCCTGGAATGCAAAACAGGTTCCGTCATTTATACAAATATACATCTATTTTAAAGAA 1639
QY 464 GlnGlyTyrAspTrpSerGlnProPheSerProGlnGlnLysGlnPheLysCysProIle 483
1630 AGCAAAATATAAACGATCCAGTGGGCTGCTCCTCAAGTATTCACAACTGATCTATC 1689
QY 484 LysGlnGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArgHisGlySerLys 503
1690 AAAAGAGAGATGCAATTTACAGTGGTGAAGGGAAGTCTTGGCGGCATGATCTAAT 1749
QY 504 IleTrpValAsnGlnGluTrpLysLeuValTyrPheGlnGlyThrLysAspThrProLeu 523
|||||.....
Db 1750 ATCCAAATTGATGAAGCAGAGGCTGGTATATTTTGAAGCAGCAAAAGACTCCCTTTA 1809
QY 524 GluHisHisLeuTyrValValSerTyrGluAlaAlaGlyGluIleValArgLeuThrThr 543
1810 GAGCATACCTGTACGTACGTACGTTACGTTAAATCTCGAGAGGTGAGCAAGGCTGACTGAC 1869
QY 544 ProGlnHisSerHisSerCysSerMetSerGlnAsnPheAspMetPheValSerHisTyr 563
1870 CGTGGCTACGACATTTCTGTCATGACGACGCTGCTCTTTTATTAAGAGTAT 1929
QY 564 SerSerValSerThrProProCysValHisValTyrLysLeuSerGlyProAspAsp 583
1930 AGTACCAAGAAATCCACACTGTGTCTCCCTTACAAAGCATCAAGTCCGGAAGATGAC 1969
QY 584 ProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaAlaSerCysProPro 603
1990 CCAACTTGCAAAACAAAGAAATTTTGGCCACCATTTTGGATTC-AGT----- 2036
QY 604 AspTyrValProProGluIlePheHisPheHisThrArgSerAspValAlaGlyLeuTyrGly 623

Db 2036 ----- 2036
QY 624 MetIleTyrLysProHisAlaLeuGlnProGlyLysHisProThrValLeuPheVal 643
Db 2036 ----- 2036
QY 644 TyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeuArg 663
2037 -----CCTCAGGTGCGAGTTGGTGAATATATCGTTTAAAGAGTCAAGTATTTCCG 2087
QY 664 LeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySerCys 683
2088 TTGAATACCTTACGCTCTCTAGGTTATGTGTGTCTAGTACGACAAAGGCGATCTCTGT 2147
QY 684 GlnArgLeuLeuArgPheGlnGluIleAlaLeuLysAsnGlnMetGlyGlnValGlnGlu 703
2148 CACCGAGGCTTAAATTTTGAAGCGCCCTTTAAATATATTAATGGGTCAAAATGAATTTGAC 2207
QY 704 AspGlnValGlnGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArg 723
2208 GATCAGGTGGAAGAGCTCCATATATCTAGCTTCTGATATGATTTTCATTTGACTTGATCGT 2267
QY 724 ValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLys 743
2268 GTGGGCAATCCACGCTGCTCTAGGATGAGATACCTCTCCCTGAGGCAATTAAATGCAAGG 2327
QY 744 ProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyrAsp 763
2328 TCAGATATCTTCACAGGTGTCTGATTTGCTGGGGCCCACTGCTGCTGATCTTCTATGAT 2387
QY 764 ThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGlyTyrGluAla 783
2388 ACAGGATATACAGGAACTTATATGTTGGTCCACCCTGACCAGCAATGAAACAGGCTATTACTTA 2447
QY 784 GlySerValAlaLeuHisValGlnLysLeuProAsnGlnLysProAsnArgLeuLeuIleLeu 803
2448 GGATCTGTGGCCATGACAGCAAGAAAGTCCCTCTGAAACCAATTCCTTACTGCTCTTA 2507
QY 804 HisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuValSerGlnLeu 823
2508 CATGTTCTCTCGATATGAGATGTCATTTGGCACATACAGATATATACGAGTTTTTTA 2567
QY 824 IleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnLysSerIleArg 843
2568 GTGAGGCTGTGAAAGCCATATGATTTACAGATCTATCTCAGGAGACACAGCATTAAGA 2627
QY 844 CysProGlnSerGlyGlnHisTyrGlnValThrLeuLeuHisPheLeuGlnGluTyrLeu 863
2628 GTTCTGTAATCGGAGAACATTTATGAACGTCATCTTTTGACATCACTTCAAGAAACCTT 2687
RESULT 13
US-09-976-674-20
; Sequence 20. Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 4676
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-20

Alignment Scores:

Pred. No.:	4,166-271	Length:	4676
Score:	2608.00	Matches:	476
Percent Similarity:	72.26%	Conservative:	131
Best Local Similarity:	56.67%	Mismatches:	181
Query Match:	56.13%	Indels:	53
DB:	10	Gaps:	3

US-09-976-674-3 (1-863) x US-09-976-674-20 (1-4676)

QY 24 PheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGlySerArgLys 43
DB 316 TTTTATGTTGAGCGGATTCCTGAGACTGACGTTAAAGCTGCTGCCATACCAAGAAA 375

QY 44 TyrSerGlyLeuIleValAlaLysAlaProHisAspPheGlnPheValGlnLysTrpAsp 63
DB 376 TATCATGGCTACATGTGGCTAAGCCACACATGATTCATGTGTTGTGAAGAGAAATGAT 435

QY 64 GluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyLysProTyrGlySerArgLys 83
DB 436 CCAGATGGACCTCATTCACAGACAAATCTATTACCTTGCCATGCTGTGGAGACAGAGAA 495

QY 84 AsnSerLeuLeuTyrSerGlyIleProLysLysValArgLysGlnAlaLeuLeuLeuLeu 103
DB 496 AATACACTGTTTATCTGGAATTCACAAACATCATAGACACAGCTTAATGCTC 555

QY 104 SerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisGlyValTyrSerArg 123
DB 556 TCTTGGAAACCTCTTTTGGATCTTTTTCAGGCAACCTGACATGTAATGTATCTCGA 615

QY 124 GluGlnGluLeuLeuArgGlnArgLysArgLeuGlyValPheGlyIleThrSerTyrAsp 143
DB 616 GAAGAAAGAACTATTAGAGAAAGAAAGCCATTGGAACAGTCGGAATGCTTCTTACGAT 675

QY 144 PheHisSerGlySerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPheHisCysArg 163
DB 676 TATCACCAGAGAAAGTGAACATTTCTGTTTCAAGCCGAGTAGAATTTATCAACGTAANA 725

QY 164 AspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIleLysThrGln 183
DB 736 GATGGAGGCCACCAAGGATTTACGCAACAACCTTTAAGGCCCAATCTAGAGGAACTAGT 795

QY 184 CysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPhe 203
DB 796 TGTCCCAACATACGATGATGATCCAAATATATGCCCTGCTGATCAGACACGTGATGCTTT 855

QY 204 IleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyLysGlnArgArgLeuThr 223
DB 856 ATTCATAGCAGACGATATTGGATATCTAACATCGTAACGAGAGAAAGAGACATCAGT 915

QY 224 PheCysHisGlnGlyLeuSerAsnValIleAspAspProLysSerAlaGlyValAlaThr 243
DB 916 TATGTCCACAAATGAGCTAGACATGAGGAAGAGATGCCGATCAGCTGAGTCCGTACC 975

QY 244 PheValIleGlnGlnLysPheAspArgPheThrGlyTyrTrpTrpCysProThrAlaSer 263
DB 976 TTTGTTCTCCAGAAATTTGATAGATATTCGGCTATTGGTGGTCCAAAAGCTGAA 1035

QY 264 TrpGlnGlySerGlnLysLysTrpLeuArgIleLeuTyrGlnLysValAspGluSer 283
DB 1036 ACAATCCCAAGTGTGCTGT--AAATTCCTAGAAATTCCTATATGAGAAATGATGATCT 1092

QY 284 GluValGlnValIleHisValProSerProAlaLeuGlnLysTrpAspSerTyr 303
DB 1093 GAGGTGGAATTTATTCATGTTTACATCCCTATCTGGAAACAGGAGGCGAGATTATCT 1152

QY 304 ArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAlaGlnPheGln 323
DB 1153 CGTTATCTTAAACAGAGTACAGCAATCTTAAAGTCTCTTTTAAAGATGTAGAAATATAG 1212

QY 324 ThrAspSerGlnGlyLysIleValSerThrGlnGlnLysGlnLeuValGlnProPheSer 343

DB 1213 ATTGATGCTGAAGGAAGATCATAGATGTCAATAGAACTAATTCACCTTTTGAG 1272

QY 344 SerLeuPheProLysValGlnTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyr 363
DB 1273 ATTCTATTTGAAGAGATGAAATATATTCGACAGGCTGGATGGATCTCTGAGGGAATATAT 1332

QY 364 AlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnIleValLeuLeuProPro 383
DB 1333 GCTTGTCATCCTACTAGATGCTGCCAGACGCGGCTACAGATATGTTGATCTCAGCT 1392

QY 384 AlaLeuPheIleProSerThrGlnAsnGlnGlnAlaArgLeuAlaSerAlaArgAlaVal 403
DB 1393 GAATTATTTATCCCAAGATGATGTTATGGAAGGACAGACATCTATTGAGTCACTG 1452

QY 404 ProArgAsnValGlnProTyrValIleTyrGlnGlnValIleThrAsnValTrpIleAsnVal 423
DB 1453 CCTGATTCGTGCGCCACTAATTAATTCATGAGAAACAAACACATCTGTGATTAATATC 1512

QY 424 HisAspIlePheTyrProPheProGlnSerGlnGlyLysAspGlyLeuLeuLysPheLeuArg 443
DB 1513 CATGACATCTTTCATGTTTTCGCCAAAGTCAAC--GAAGAGAAATTCAGTTATTTT 1569

QY 444 AlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSer 463
DB 1570 GCCTTGAATGCAAAACAGGTTCCGTCATTTTACAAATATTCATCTATTTTAAAGGAA 1629

QY 464 GlnGlyTyrAspTrpSerGlnProPheSerProGlyGlnAspGlnPheLysCysProIle 483
DB 1630 AGCAAAATATAAAGATCCAGTGTGGGCTGCCGTCGCAAGGATTTCAAGTCTCTATC 1689

QY 484 LysGlnGlnIleAlaLeuThrSerGlyGlnTrpGlnValIleuAlaArgHisGlySerLys 503
DB 1690 AAAGAGAGATGACATTTTACACAGTGTGAATGGAAATCTTGTGGCGCATGATTAAT 1749

QY 504 IleTrpValAsnGlnLysThrLysLeuValTyrPheGlnIleLysAspThrProLeu 523
DB 1750 ATCCAGTGTGATGATGACGAAAGGCTGGTATTTTGAAGGCAACCAAGCTCCCTTAA 1809

QY 524 GlnHisHisLeuTyrValValAspTyrGlnAlaIleGlyLysIleValArgLeuThrTrp 543
DB 1810 GACATCACCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1869

QY 544 ProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspPheValSerHisTyr 563
DB 1870 CGTGGCTACGACATCTTCTGCTGATCAGTACAGACATGATCTTTTAAATAGTAT 1929

QY 564 SerSerValSerThrProProCysValHisValTyrLysLeuSerGlyProAspAspAsp 583
DB 1930 AGTAACCCAGAAAGATCCACACTGTGTGCTGCTTAAAGCTATCAAGTCTGTAAGATGAC 1989

QY 584 ProLeuHisLysGlnProArgPheThrAlaSerMetMetGlnAlaIleSerCysProPro 603
DB 1990 CCAACTTGCAAAACAAAGGAATTTTGGCCACCAATTTTGGATTCACAGGCTCTTCTCT 2049

QY 604 AspTyrValProProGlnIlePheHisPheHisThrArgSerAspValArgLeuTyrGly 623
DB 2050 GACTATACCTCCGACAAATTTCTCTTTTGAAGAGTACCTGATTTACATTTGATGGG 2109

QY 624 MetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValIlePheVal 643
DB 2110 ATGCTCTACAAAGCTCATGATCTACAGCCGTGGAAGAAATATCTCATGCTGCTCATTA 2169

QY 644 TyrGlnGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeuArg 663
DB 2170 TATGTGTGCTG----- 2180

QY 664 LeuAsnThrLeuAlaSerLeuGlnTyrAlaValAlaValIleAspGlyArgGlySerCys 683
DB 2180 ----- 2180

QY 684 GlnArgGlyLeuArgPheGlnGlnValAlaLeuLysAsnGlnMetGlyGlnValGlnIleGly 703
DB 2181 -----GTCGCAAAATGAGAAATTTGAC 2198

Oy	704	AspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArg	723
Db	2199	GATCAGGTGGGAAGACATCCAAATCTACAGCTTCCTACATGATGATTTCTATCTTCAATGATCGT	2258
Oy	724	ValAlaIleHisGlyTyrPseTyrGlyGlyPheLeuSerLeuMetCylLeuIleHisLys	743
Db	2259	GTGGGCATCCACGGCTGTGCTCTATGGAGAGATACCTCTCCCTCATGGCATTAATGCAGAGG	2318
Oy	744	ProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTyrMetAlaTyrAsp	763
Db	2319	TCAAGATATCTTCAGAGGTTCGTATATGGCGGGCCCCAGTACACTGTGGATCTTCATGAT	2378
Oy	764	ThrGlyTyrThrGluTAspTyrMetLeaPvalProGluAsnAspGlnHisGlyTyrGluAla	783
Db	2379	ACAGGATATACGGAAACCTTATATGGCTACCCCTGCACCAATGAACAGGGCTATTAAGCTTA	2438
Oy	784	GlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeuIleLeu	803
Db	2439	GGATCTGTGGCCATGCACAGCAAGAAAGATTCCCTCTGCACCAAAATGGTTTACTGCTCTTA	2498
Oy	804	HisGlyPheLeuAspArgLysuAsnValHisPhePheHisThrAsnPheLeuValSerGln	823
Db	2499	CATGGTTCCTCGGATGGAAATGTCCATTGGCACATACCAAGTATATTAATCTGAGTCTTTTAA	2558
Oy	824	IleAsnAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluAsnHisSerIleArg	843
Db	2559	GTGAGGGCGGTGAAGACCATATGATTACACAGATCTATCTCAGAGAGACACAGCATAGA	2618
Oy	844	CysProGluSerGlyLysHisTyrGluValThrLeuLeuHisPheLeuGlnGluTyrLeu	863
Db	2619	GTTCCTGAATCGGAGAAACATTTATGAAGATGCATCTTTCTACCTACCTTCACAAAAACCTT	2678

```

RESULT 14
US-09-976-674-26
: Sequence 26, Application US/09976674
: Patent No. US20020115843A1
: GENERAL INFORMATION:
: APPLICANT: Qi, Steve
: APPLICANT: Akinsanya, Karen
: APPLICANT: Riviere, Pierie
: APPLICANT: Junien, Jean-Louis
: TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
: FILE REFERENCE: 70669
: CURRENT APPLICATION NUMBER: US/09/976,674
: CURRENT FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: US 60/240,117
: PRIOR FILING DATE: 2000-10-12
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 26
: LENGTH: 2411
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-976-674-26

Alignment Scores:
Pred. No.:          4e-266
Score:              2576.00
Percent Similarity: 83.39%
Best Local Similarity: 81.13%
Query Match:        55.45%
DB:                  10
                    10
Gaps:                6
                    6
Length:             2411
Matches:            503
Conservative:       14
Mismatch:           54
Indels:              49
                    6

```

QY	1	MeLaIaThrThrglyThrProHnraIaAspaArgGlyAspAlaIaIaLaThraAspAspPro	20
Db	367	ATGGCCACCCACCGGACCCCAAGCGCCGACCGAGCGAGCAGCGCCACAGATGACCCG	426
QY	21	AlaIaIaArgPheGlnValGlnIySHSSerTrrPaspGlyLeuArgSerIleIleHisGly	40
Db	427	GCCGCCGCTTCACGGGCGAAGCACTGTGGGAAGGGCTCCGAGGATCATCCACGGC	486

OY	41	SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln	60
Db	487	AGCCGCAAGTACTCGGGCCCTCATTTGTCACAAAGGCGCCACAGACTTCCAGTTTGTCAG	546
OY	61	LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrIleuLysIleMetProTyrGly	80
Db	547	AAGAGGATGTAGTGTGGGCCCCCATCTCCACGGCTCTACTACCTGGGAATGGCATATGGGC	606
OY	81	SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu	100
Db	607	AGCCGAGAGACTCCCTCCTCTACTGATTCACCAAGAGGTCCGGAAAGAGGCTCTCG	666
OY	101	LeuLeuLeuSerTrrPlySGlnMetLeuAspHisPheGlnAlaThrProHisIleGlyVal	120
Db	667	CTGCTCCCTGTCCTCGAAGCAGATGCTGGATCATTTTCAGAGCCACGCCCATCATGGGGTTC	726
OY	121	TyrSerArgGluGluGluLeuLeuArgGluAlaArgLysArgLeuGluValaPheGlyIleThr	140
Db	727	TACTCTCGGAGAGAGAGACTGTCTGAGGAGCGGAAGCCCTGGGGGTCTTCGGCATACC	786
OY	141	SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe	160
Db	787	TCTTACGACTTCCACAGCGAGAGTGGCTTCTCTCTCCAGGCCAGCAAGACCTCTTC	846
OY	161	HisCysArgAspArgGlyLysLysAsnGlyPheMetValSerProMetLysProIleuGluIle	180
Db	847	CACGTCCCGCGAGCGGCGCAAGAACGGCTTATGTGTCCCTATGAACCGGTGGAAATC	906
OY	181	LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe	200
Db	907	AAGACCCAGTCTCAGGGGCCCGGATGTGACCCCAAAATCTGCCCTCCGACCTGCTTC	966
OY	201	PheSerPheIleAsnAsnSerAspLeuTrrPValAlaAsnIleGluThrGlyGluGluArg	220
Db	967	TTTCTCTTCACTCAATTAACAGCGACTGTGGGTGCCCAACTCGAGCAGCGAGAGCGG	1026
OY	221	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly	240
Db	1027	CGCGTACCTTCTCCACCAAGGTTTATCCAAATGCTGTGATGACCCCAAGTCTCGGGGT	1086
OY	241	ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTrrPrrCysPro	260
Db	1087	GTGGCCACTTTCGTCATACAGAGAAATTTGACCGCTTCACGTGTACTGCTGGTGGCC	1146
OY	261	ThrAlaSerTrrGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluVal	280
Db	1147	ACACGCTCTCTGGAAAGGTTTCAGAGGGCTCAAGCCTCGAATCCTGTATAGGAAGTC	1206
OY	281	AspGluSerGluValGluValIleHisValProSerProAlaIleuGluLysIleArgLysThr	300
Db	1207	GATAGTCCCGAGGGTGAGGTCATTCACTGCCCTCTCGCTTACAAAGAAAGAACG	1266
OY	301	AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla	320
Db	1267	GACTCGTATGSGTACCCAGAGACAGGACAGCAAAATCCCAAGATTGCTTGAACCTGGCT	1326
OY	321	GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln	340
Db	1327	GAGTTCCAGACTGACACCCAGGCGAGATCTGTCCACCCAGAGAGAAAGCTGTGTGAG	1386
OY	341	ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrrPrrArgAsp	360
Db	1387	CCCTTCACTGCTGCTTCCGAAAGTGGATACATGCCACAGGGCGGGGTGGACCCGGGAT	1446
OY	361	GlyLysTyrAlaTrrPalaMetPheLeuAspArgProGlnGlnTrrPheuGlnLeuValLeu	380
Db	1447	GGCAATATACGCTGTGGCAATGTTCTGTGACCGGCCACAGATGGGTCTCCAGTCTCTC	1506
OY	381	LeuProProAlaLeuPheIleProSerThrGluAsnGluGluArgLeuAlaSerAla	400
Db	1507	CTTCCCCCGGCGCTTCACTCCGAGCACAGAAATGAGAGACAGGCGCTTACCTCTCC	1566

```

QY 401 ArgAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTrp 420
Db 1567 AGAGTGTCCCGAGGAAGTCCAGCCGATATGTGTAGAGAGAGTCCCAAGCTCTGG 1626
QY 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluAspGluLeuGly 440
Db 1627 ATCAATGTCATGACATGCTTCTATCCCTCCCAATCAGAGAGAGAGAGAGAGCTCTGC 1686
QY 441 PheLeuArgAlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValAlaVal 460
Db 1687 TTTCTCCCGCCCAATGAATGAAGAACCGGCTTCTCCATTTGTACAAATGTACCGCGCTT 1746
QY 461 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGluGluAspLupheLys 480
Db 1747 TTAATAATCCAGAGGCTACGATTTGAGTGAAGCCCTTCAGCCCGGGGAAGGTAGCAGAGC 1806
QY 481 -----CysPro 482
Db 1807 CTGACGAATGCTGTGACATCGCGGTAGTACGTTGCTTCAMATGCTGTTGTCTCA 1866
QY 483 Ile-----LysGluGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArg 499
Db 1867 TTGGTGGGCCCCCAGCTACGACACACCCCTGGGGGAAGGAACAGAGGATCGGCAGG 1926
QY 500 HisGlySer-----LysIleTrpValAsnGluGlu-ThrLysLeuValTyr 514
Db 1927 AAGCCAGCTTCCCGCAGTACTGCATGATGTGCGAGGGCTTAGAGCACCACCTGTGGC 1986
QY 514 rPheGlnGlyThrLysAspTrpProLeuGluHisLeuTyrValValSerTyrGluAl 534
Db 1987 TTATTACAGCAGCAGATTTACT-----GAGCACCTCCCTGNGCCAGG----- 2029
QY 534 AlaIleGluIleValAlaGluLeuThrThrProGlyPheSerHisSerCysSerMetSerI 554
Db 2030 -----CCCTTAGCACACAGGAGGTTGGCCACCTACGAGCCACAGCTA 2073
QY 554 AsnAspPheAspMetPheValSerHisTyrSerSerValSerThrProCysValHisVa 574
Db 2074 A-----ATCCGCCACACCACTGTGTTGATMA 2100
QY 574 LTrLysLeuSerGlyProAspAspProLeuHisLysGlnProArgPheTrpAla 593
Db 2101 ATAAAGTTTATTATGGCACTGAGCCACAGCCACTGTTTACAGAGACTGTGTGGTGC 2158

RESULT 15
US-10-044-090-843
; Sequence 843, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 843
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc. feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 2643110CB1
US-10-044-090-843

```

Alignment Scores:

```

Pred. No.: 3,32e-265 Length: 2079
Score: 2548.00 Matches: 497
Percent Similarity: 84.67% Conservative: 11
Best Local Similarity: 82.83% Mismatches: 43
Query Match: 54.84% Indels: 49
Db: 12 Gaps: 6

```

```

US-09-976-674-3 (1-863) x US-10-044-090-843 (1-2079)
QY 1 MetaIatThrGlyThrProThrAlaAspArgGlyAspAlaAlaAlaThrAspAspPro 20
Db 330 ATGGCCACACCGGAGACCCCAAGCGCCGATCGAGCAGCGAGCGCCACATATACCCG 389
QY 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
Db 390 GCCGCCCGCTTCAGGTGAGAGACACTGTGTGGAGCGGCTCCGAGCATCTCCACAGCG 449
QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 450 ACCCCCAAGTACTCGGGCTCATTTGCAACAGAGCGCCACAGCATCTCCAGTTGTGGCAG 509
QY 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGluYmetProTyrGly 80
Db 510 AAGAGGAGATGTGGGGCCCACTCCACCCCTCTTACTGAGGAATGGCATATGGC 569
QY 81 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
Db 570 AGCCGAGAGACTCCCTCTTACTGTGATTTCCCAAGAGTCCGGAAGAGCTCTG 629
QY 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyAl 120
Db 630 CTGCTCCTGCTCGGAAGCAGATGCTGATTCATTTCCAGAGCCACGCCCCACCATGGGCTC 689
QY 121 TyrSerArgGluGluGluLeuLeuArgGluAlaGlyLysArgLeuGlyAlaPheGlyIleThr 140
Db 690 TACTCTCGGAGAGAGAGTGTGTGAGGAGCGGAACGCTGGGGTCTTCGGCAACAC 749
QY 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
Db 750 TCTTAGACTTCCACAGCAGAGAGTGGCTTCTTCTTCCAGGCGCAACAGCTCTTC 809
QY 161 HisCysArgAspGlyLysLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
Db 810 CACTGGCGGAGAGCGCGGAAGAACGGCTTCATGTGTCCCTATGAAACCGCTGGAATTC 869
QY 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProLysProAlaPhe 200
Db 870 AAGACCCAGTGTCTCAGAGGCGCCGAGATGAGCCCAAAATGTGCCCTGCCACCTGACTTC 929
QY 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrLysGluGluArg 220
Db 930 TTCTCCTTATCATATACAGCAGCACTGTGGTGGCCAAATCGAGACAGCGAGAGCGG 989
QY 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240
Db 990 CGGCTGACCTTGGCCACCAAGTTTATCCAAATGCTCGGATGACCCCAAGTCTGGGGGT 1049
QY 241 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpTrpCysPro 260
Db 1050 GTGGCCACCTTGTCTATACAGGAAGATTGACAGCGCTTACGTGGTACGTGGTGTCCCC 1109
QY 261 ThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluVal 280
Db 1110 ACAGCTCTCTGGAGAGTTTCAGAGGCTTCAGAGCCTCGGAATCTGTATGAGAGATC 1169
QY 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluValArgLysThr 300
Db 1170 GATGAGTCCGAGGTGAGGTATTCACGTCCCTCTCGCGCTAGAGAAAGAGAGAGC 1229
QY 301 AspSerTyrArgTyrProArgThrGlySerLysAspProLysIleAlaLeuLysLeuAla 320
Db 1230 GACTCGTATCGGTACCCAGGAGCAGCACCACCAATCCCAAGATTCCTTGAAACGTGGCT 1289
QY 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340
Db 1290 GAGTTCCAGACTGACAGCAGGAGGATGCTGTCCAGCCAGAGAGAGAGAGCTGTGGCAG 1349
QY 341 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrArgAsp 360
Db 1350 CCTTACGCTCGCTGTCCGAGAGGTGAGTACATGCGCAGGCGCGGTGAGACCCGGAT 1409

```


[illegible]

Db	620	ptmProIleIysValAlaIcylProLeuThrArgValLeuAsnValAlaGluIleHisAs	640
QY	1756	CGAC-----CCCTGCACAGACAGCCCGCTCTGGCTACAGATGAGGAGC	1806
Db	640	paspSerIleLeuIleuLeuThrIlysAspGluIysPheIysGluIysIle-----	656
QY	1807	CAGCTGCCCCCGATATATGCTTCCACAGATCTTCATTTCCACACGCC-----	1858
Db	657	-----LysAsnIysPheLeuPro--IleThrSerTyrLysThrMetValLeuAs	672
QY	1859	-TCGATGTGGCGGTCTACGACATGATCTCAAGACCCCGCTTCGACCGAGGA	1917
Db	672	paspGlyValGluIleAsnTyrIleGluIleIleProAlaAsnLeuAsnProLysLys	692
QY	1918	GCACCCACCGCTCTCTTTATATGATGAGAGCCCGCAGTGCAGCTGTGAATACCTCT	1977
Db	692	styrProIleLeuValAsnIleTyrGlyProGlySerGln-----	706
QY	1978	CAAAAGCATCAAGTCTGGCGCTCAACACACTGGCTCCCTGGCGTACGCCGCTGT	2037
Db	707	-----ThrPheThrTyrSerSerIleuAlaPheGluGlnAlaIaVa	720
QY	2038	GATTGACGCG-----AGGGCGCTCGTGCACCGAG	2067
Db	720	IValSerGlyLeuAspValIleValLeuGlnIleGluProAlaGlyThrGlyLysG	740
QY	2068	GCTTGGTTCGAGGGGCGCTGAAAAACCAATGGCGCAGTGGAGATCGAGACCAGT	2127
Db	740	yTyrSerPheArgSerTyrPalaArgGluIysLeuGlyTyrTrpGluProArgAsp	760
QY	2128	GGAG---GGCGTGCAGTTCGCGGGCCGAGAAAGTATGGCTTCACAGCCGAGCGAGTGC	2184
Db	760	rgluValThrLysLysPheIleGlnArgAsnSerGlnHisIleAspGluSerLysIleAl	780
QY	2185	CATCCATGGCGTGCCTACACGGGGGCTTCTCTGCTCATAGGGGCTA--ATCCACAA	2241
Db	780	alIleTyrGlyPheTyrGlyIlePheThrSerLeuLysThrValGluLeuAspAsnG	800
QY	2242	CCAGGTGTCAAGTGGCCATCCGGGGTGGCCCGGTACACCGCTGGATGGCTACGAC	2301
Db	800	yAspThrPheLysTyrTrpIleMetAlaValAlaProValTrpAsnTyrThrLeuTyrAsp	820
QY	2302	AGGACATCAGGCGCTACACTGTGACGCGCTTCAGACACACGAGCGGTATGAGCGGG	2361
Db	820	rValTyrThrGluArgTyrIleMetAsnGlnProSerGluAsnHisGluGlyTyrPheGlu	840
QY	2362	TTCCGTGGCCCTGCACGCTGAGAAGCTGCCAATGAGCCCAACCGCTTGTATTCCTCA	2421
Db	840	IserThrIleGlnAsnPheLysSerPhe--GluSerLeuLysThrGluPheIleVal	859
QY	2422	CGGCTTCTGACGCAAAACGTGCACCTTTTCCACACAAATTCCTCTGCTCCCATAT	2481
Db	859	sgIyThrPheAspAsnValHisIleGlnAsnThrPheArgLeuValAspGlnLeuAs	879
QY	2482	CCGAGCAGAGG---AAACCTTACAGACTCCAGTATCCACCCCAAGAGACACAGTATTCG	2538
Db	879	nIleuLeuGlyLeuThrAsnTyrAspMetHisIlePheProAspSerAspHisSerIleAr	899
QY	2539	C 2539	
Db	899	g 899	

Search completed: December 12, 2002, 11:53:34
Job time : 83.5 secs

J. Cell Biol. 108, 1363-1373, 1989
 A:Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an in
 A:Reference number: A30107; MUID:89174971; PMID:2647766
 A:Accession: A30107
 A:Molecule type: DNA
 A:Residues: 1-82, 'H', 84-124, 'N', 126-181, 'LRLET', 189-199, 'N', 201-365, 'DFRKGRKR', 376-57
 A:Cross-references: EMBL:X15484
 A:Note: the authors translated the codon ACC for residue 572 as Asn
 C:Genetics:
 A:Gene: SGD:DAP2; SRE13; MIPS:YHR028c
 A:Cross-references: SGD:S0001070; MIPS:YHR028c
 A:Map position: 8R
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast vacu
 F:30-45/Domain: transmembrane #status predicted <TMM>
 F:63,79,110,139,392,421/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.:	4,24e-18	Length:	818
Score:	432.50	Matches:	164
Percent Similarity:	39.47%	Conservative:	91
Best Local Similarity:	25.39%	Mismatches:	262
Query Match:	9.02%	Indels:	129
DB:	1	Gaps:	28

US-09-976-674-4 (1-2617) x A30107 (1-818)

```

OY 728 GGGTGGGCCCTTCGTCATACAGAAAG---TTGAGCCGTCACGTGGTACTGGTG 784
DB 234 GLLysPrAspIrrValTyrGluGluValaPheGluAspIysAlaIaIrrPTT 253
OY 785 TGCCCAACACCTCTCGGAAGGTTCAGAGGCTCAAGACGCTGCATCTGTATGAG 844
DB 254 SerProThrGlyAspTyr-----LeuAlaPheLeu 263
OY 845 GAAGTGTAGTCCGAGGTGAGGTCATTCACGTCCTCTCTCGGTAGAAAG 904
DB 264 LysIleAspIleSerGluValaGlyGluPheIleProTyrTyrValaGluAspGluLys 283
OY 905 -----AAGACGGACGTGTACCCGAGGATCCGTCAGCAAGAAATCCCAAG 952
DB 284 AspIleTyrProGluMetAlaGserIleLysTyrProLysSerGlyThrProAsnProHis 303
OY 953 ATTGCCTTGAACGTGGCTGATTCACAGACTACAGCCAGGCAAGATGTCGACCCAG 1012
DB 304 AlaGluLeuTyrValaTyrSerMetLysAspGly----- 314
OY 1013 GAAAGAGAGCTGGTGCAGCCCTCAGCTGCTGCCGAAGTGGAG----- 1060
DB 315 -----ThSerPheHisProArgIleSerGlyAsnLysLys 326
OY 1061 -----TACATCCGACAGGCGGGGTGACCCGGATGGAATACGCTGGGCC 1108
DB 327 AspGlySerLeuLeuIleThrGluValaThrTyrValaGlyAsnGlyAsn-----ValLeuVal 345
OY 1109 ATGTCTTCGACCGGCCGACAGAGTGGCTCCAGCTGCTCTCCCTCCCGGCCCTGTTG 1168
DB 346 LysThrThrAspArgSerSerAspIleLeuThrValaPheLeuIleAspThr----- 362
OY 1169 ATGCCGACGACAGAGATGAGGACACGGGCTAGCCTTCGACAGAGTGTGCCAGAAAT 1228
DB 363 IleAlaLysThrSerAsn----- 368
OY 1229 GTCCAGCCGTATGTGTGTAGAGAGAGTACCAAC-----GTCTGATCAATGTTGAT 1282
DB 369 -----ValValaLysAsnGluSerSerAsnGlyLysTyrTyrPrgIuIleThrHis 384
OY 1283 GACATCTTATCCCTTCCCAATCAGAGGAGAGACAGAGCTGTGCTTCTCCGGGCC 1342
DB 385 AsnThrLeu-----PheIleProAla 391
OY 1343 AATGAATGC-----AAGACGGCTTCGTCATTTGTACAAAGTACACCGCGTT 1390
DB 1343 AATGAATGC-----AAGACGGCTTCGTCATTTGTACAAAGTACACCGCGTT 1390
  
```

```

DB 392 AsnGluThrPheAspArgProHisAsnGlyTyr-----ValAspIle 405
OY 1391 TTAAATATCCAGGCGTACAGTTGAGTGAAGCCCTTC---AGCCCGGGGAATGAAATT 1447
DB 406 LeuProIleGlyLysTyrAsnHisLeuAlaTyrPheGluAsnSerAsnSerHisTyr 425
OY 1448 AAGTGGCCCATTAAGAAAGATTTGCTGTGACAGCGGTGAATGGAGAGTTTGGCGAGG 1507
DB 426 Lys-----ThrLeuThrGluGlyLysTyrPrgIuVala----- 436
OY 1508 CAGCGCTCAAGATCTGGGTCATGAGAGACCAACAGCTGTGACTTCCAGGCGACCAAG 1567
DB 437 AsnGlyProLeuAlaPheAspSerMetGlyAsnArgLeu---TyrPheIleSerThrArg 455
OY 1568 GACAGCGCGGTGAGACACCACTCTGAGTGTGATGAGCGCGCGCGAGATCGTA 1627
DB 456 LysSerSerThrGluArgHisValTyrTyrIleAspLeuArgSerProAsnGluIleIle 475
OY 1628 CGCCTCACACGCGCCGCGC-----TTCTCCATAGCTGTCCATGAGCCAGAAC 1675
DB 476 GluValThrAspThrSerGluAspGlyValTyrAspValSerPheSerSerGlyArgArg 495
OY 1676 TTGCATATGTTCTGACACCACTACAGCAGCGTGAACAGCGCCGCTGCGTG-----CAC 1729
DB 496 PheGlyLeuLeuThrTyrTyrGlyGlyProLysValProTyrGluLysIleValAspPheHis 515
OY 1730 GTCTACAGCTGAGCGGCGCCGAGCAGACCCCTGACACAGCGCCGCTTGGGCT 1789
DB 516 SerArgLysAlaGluLysCysAspLysGlyAsnValLeuGlyLysSerLeuTyrHisLeu 535
OY 1790 AGCATGTGAGAGCA---GCCACGTGCGCCCGGATTAAT---GTCTCCACAGATCTTC 1843
DB 536 GluLysAsnGluValLeuThrLysIleLeuGluAspTyrAlaValProArgLysSerPhe 555
OY 1844 -----CATTTCCACACGCGCTCGATGCGCGCTCTACGGC 1879
DB 556 ArgGluLeuAsnLeuGlyLysAspGluPheGlyLysAspIleLeuValAsnSerTyrGlu 575
OY 1880 ATGATCTACAAAGCCCGCAGCTTCGACCGGAGGAAGAACACCCGCTCCTTGTA 1939
DB 576 IleLeuProAsnAspPheAspArgIuThrLeuSerAspHisTyrProValPhePhePheAla 595
OY 1940 TATGAGGCGCCGAGGTGACGCTGTGTAATACCTCTTCAA---GGCATCAAGTACTTG 1996
DB 596 TyrGlyGlyProAsnSerGlnGluValaValLysThrPheSerValaGlyPheAsnGluVal 615
OY 1997 CGGCTCAACACACTGCGCTCCCTCGGCGGTACCGCGCTGTTGATTTGACGACGAGGCTCC 2056
DB 616 -----ValAlaSerGlnLeuAsnAlaIleValaValaValaLysPrgIuArgLysThr 632
OY 2057 TGTACAGGAGGCGTTCGTTGCGAAGGGCCCTGAATAACCAATGGGCGAGTGAAGATC 2116
DB 633 GlyPheLysGlyGlnAspPheArgSerLeuValaArgAspArgLeuGlyAspTyrGluAla 652
OY 2117 GAGGACCAAGGTGAGGCGCTGACGTTCTGGCCGAGAAATATGAGC-----TTGATC 2167
DB 653 ArgAspGlnIleSer-----AlaAlaSerLeuTyrGlySerLeuThrPheVal 668
OY 2168 GACCTGAGCGGATGTCATCCATGCTGTGCTACAGGGGCGTTCCTGCTGCATAGGGG 2227
DB 669 AspProGlnLysIleSerLeuPheGlyTyrSerTyrIleLysIleLeuThrLysThr 688
OY 2228 CTAATCCACAAG---CCCCAGGTGTTCAAGGTGGCCATCCCGGGTGGCCGTCACCGTTC 2284
DB 689 LeuGluLysAspGlyGlyAlaGlnHisPheLysTyrGlyMetSerValaAlaProValaThrAsp 708
OY 2285 TGATGGCTTACGACACAGGCTACGACGCTACGACGCTACGACGCTGACGACGACGACG 2344
DB 709 TyrArgPheTyrAspSerValTyrThrGluArgTyrMetHisThrProGlnGluAsnPhe 728
OY 2345 CACGGCTATGAGGCGGTTCCGCGGCGCTGCACGTGAGAGAGCGCCCATAGGCCCAAC 2404
DB 729 AspGlyTyrValaGluSerSerVal-----HisAsnValThrAlaLeuAlaGlnAlaAsn 746
  
```

Db 227 SerAspTyrLeuTyrGlu-----GluGluIle 235
QY 758 TTGCACGGCTTCACGGGTACTGGTGGCCACAGCGCTCTGGGAGAGTTCCAGAGGC 817
Db 236 LeuLysThrHisIleAlaHisThrSerProAspGlyThr----- 249
QY 818 CTCAGACGGCTGCAATCTGTATAGAGAACTGCATGATCCGAGATGGAGTCAATTCAC 877
Db 250 -----ArgLeuAlaTyrAlaThrIleAsnAspSerArgValProValMetCys 265
QY 878 GTCCCTCTCTCGCGCTAGAAAGAAAGAGACG---GACTGTATCGGTACCCAGAGCA 934
Db 266 LeuProThrTyrThrGlySerValTyrProThrAlaLysProGlyHisTyrProLysAla 285
QY 935 GGCAGCAAGATCCCAAGATTCGCTGAACCTGCTGATGCCAGACACAGCAGCAGGC 994
Db 286 GlyLysGluAsnProSerIleSerLeuHisValIle-----Gly 298
QY 995 AAGATCGTCTCGACCCAGAGAGAGAGAGCTGGTGCAGCCCTTCAGCTCGCTGGCCGAG 1054
Db 299 LeuAsnGlyProThrHisAspLeuGluMetThrProProAspAsp-----ProArg 315
QY 1055 GTGGAG-----TACATCCGACGAGCGCGGTGCAGCCGGATGCCAAATACCGCTGGCC 1108
Db 316 MetArgGluTyrTyrIleThrMetValLysTyrPalaThrSerThrLysValAla---Val 334
QY 1109 ATGTTCCTGGACGGCCCGCCACAGTGGCTCAGCTCGTCTCTCCCGCCGCTGTCC 1168
Db 335 AsnThrLeuSerArgAlaGlnAsnValSerIleLeuThrLysCysAspAlaThrThrGly 354
QY 1169 ATCCGACGACAGAAATGAGAGACAGCGGCTAGCTCTGCGACAGCTGCCACAGANT 1228
Db 355 ValCysThrLysLysHisGluAspGlu----- 363
QY 1229 GTCCAGCCGATGTGGTGTACGAGAGGTCAACAGCTGGAGTC-----AAGCTT 1279
Db 364 -----SerGluAlaThrLeuHisArgGluAsnGlu 373
QY 1280 CATGACATCTTCATCCCTTCCCGCCATCAGAGGAGAGAGAGAGCTGGTTCCTCCGC 1339
Db 374 GluProValPhe-----SerLysAspGlyArgLysPhePhePheValArg 388
QY 1340 GCCAATGATGCAAGACCGGCTTCTGCATTTGACAAAGTCAACCGCGCTTTAAATCC 1399
Db 389 AlaIlePro---GlnGlyGlnGlnGlyLysPheThrHisIleThrVal----- 403
QY 1400 CAGGGCTACGATTGGAGTACCGCTTCAACCGCGGAGAAATGAATTAATGCCCCCAT 1459
Db 404 -----SerSerSerGlnProAsnSerSerAsnAspAsn-----Ile 415
QY 1460 AAGGAAGAGATTCCTGACGAGCGGGAATGGAGGTTTGGGAGGACAGCGCTCCAG 1519
Db 416 Gln-----SerIleThrSerGlyAspThrAspVal-----ThrLys 427
QY 1520 ATCTGG---GTCAATGAGAGAACCAAGCTGTACTTCCAGGGCACCAGACAGCGCG 1576
Db 428 IleLeuSerTyrAspGlnLysArgSerGlnIleTyrPheLeuSerThrGluAspLeuPro 447
QY 1577 CTGGAGACACCACTCTACGTGGTCACTANTAGCGCGCGCGGAGANTGTACGCTCACC 1636
Db 448 ArgArgArgGlnLeuTyrSerAlaSer-----Thr 457
QY 1637 ACGCGCGGCTTCCTCCCATAGCTGCTCATGAGCCAGAACTTCGACATCTTGTACGCCAC 1696
Db 458 ValGlySerPheAsnArgGlnCysLeuSerCysAspLeuValAspAsnCysThrTyrPhe 477
QY 1697 TACAGCAGCGTACGACGCGCGCTGCGTACCTTACAGCTAGAGGAGCGCCGAC--- 1753
Db 478 SerLaserPheSerProGlyAlaAspPheLeuLeuLysCysGlnGlyProGlyVal 497
QY 1754 -----GACGACCGCCCTGCACAGACAGCGCGCGCTTCTGGGCTAGCATGATG 1798
Db 498 ProThrValSerValHisAsnThrThrAspLysLysMetPheAspLeuGluThrAsn 517

QY 1799 GAG---GCAGCGAGCTGCCCCCGGATATGTCCTCCAGAGATC---TTCATTTTCCAC 1852
Db 518 GlnHisValGlnLysAlaIleSerAspArgGlnMetProLysValGluTyrArgLysIle 537
QY 1853 ACGCGCTCGATGTCCGCGCTACGCGCATGTACAGAGCCCGCCAGCTTCCAGCGAGG 1912
Db 538 GluThrAspAspTyrAsnIleProIleGlnIleLeuLysProIleThrPheThrAspThr 557
QY 1913 AAGAGACACCCCGCTCTTTGTATATGAGAGGCCCCAGAGTGCAGCTGTGATATAC 1972
Db 558 AlaHisTyrProLeuLeuLeuValValAspGlyThrProGlySerGlnSerValAlaGlu 577
QY 1973 TCCCTCAAGGACATCACTACTGCGGCTCAACACAGCGGCTCCGCTGAGCCGCTG 2032
Db 578 LysPhe---AlaValThrTrpGluThrVal---MetValSerSerHisGlyAlaVal 595
QY 2033 GTTGATGTACGCGACGAGCGCTCTGTACAGAGGCTTCCGTTCCAGAGGCGCTGAAA 2092
Db 596 ValLysCysAspArgLysArgGlySerGlyPheGlnGlyThrArgLeuLeuHisGluValArg 615
QY 2093 AACCAATGGCCCAAGTGCAGATCGAGACAGACAGTGAAGGCGCTGCAGTTGGCCGAG 2152
Db 616 ArgArgLeuGlySerLeuGlnGluLysAspGlnMetGluAlaValArg---ValMetLeu 634
QY 2153 AAGTATGGCTCATACAGCTGAGCGAGTGCAGTCCATTCATGCTGCTGCTCAGCGGCTTC 2212
Db 635 LysGluProTyrIleAspLysThrArgValAlaValPheGlyLysAspTyrGlyTyr 654
QY 2213 CTCTCGCTCATACGCGCTTATATCCAGAG-----CCSCAGTGTTCAGAGTGGCC 2260
Db 655 LeuSerThrTyrLeuLeuLeuProAlaLysGlyAspGlnAlaProValPheSerCysGly 674
QY 2261 ATCGCGGTGCGCGGTACCGCTGCGATGCGCTTACACAGAGTGAAGTGAAGCTTAC 2320
Db 675 SerAlaLeuSerProIleThrAspPheLysLeuTyrAlaSerAlaPheSerGluArgTyr 694
QY 2321 ATGGACGCTCCGAGAACACACAGACAGCGCTATGAGCGCGGCTCCGCTGCGACG 2380
Db 695 LeuGlyLeuHisGlyLeuAspAsnArgAlaTyrGluMetAlaLysValAlaHisArgVal 714
QY 2381 GAGAGCTGCCCAATGACGCCCGCTGTACTTCTCCAGCGCTTCCAGGAGAAAC 2440
Db 715 SerAlaLeu-----GlnGlyGlnGlnPheLeuValIleHisAlaThrAlaAspGluLys 732
QY 2441 GTGCACTTTTCCACACAACTCTGCTGCTCCCACTGATCCGACGAGGAAACCTTAC 2500
Db 733 IleHisPheGlnHisThrAlaGluLeuIleThrGlnLeuIleLysGlyLysAlaAsnTyr 752
QY 2501 CAGCTCCGATCTACCCCAACAGAGACAGATGATTCGCTGCGCGGAGTCCGCGGAGCAG 2560
Db 753 SerLeuGlnIleTyrProAspGluSerHisTyrPheSerSerAlaAlaLeuGlnIleHis 772
QY 2561 TATGAGTACAGTGTCTGCTACTTCTACAGGAA 2593
Db 773 LeuHisArgSerIleLeuGlyPhePheValGlu 783
RESULT 14
A30107
dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YHR028C
C:Species: Saccharomyces cerevisiae
C:Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 23-Mar-2001
C:Accession: S46780; A30107
R:Du, Z.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 8082.
A:Reference number: S46773
A:Accession: S46780
A:Molecule type: DNA
A:Residues: 1-818 <DDZ>
A:Cross-references: EMBL:U10399; NID:9500689; PIDD:AA68879.1; PID:9500698; GSPDB:GNO
R:Roberts, C.J.; Pohlig, G.; Rothman, J.H.; Stevens, T.H.


```

Db 350 AspaLaThrThnGlyValLysThrLysHisGluAspGlu----- 363
Oy 1214 GGTGTCACAGAGATGTCAGCGGTATGTGTAGCAGAGAGTCCACAACTGTGATC 1273
Db 364 -----SerGluAlaThrPhe----- 368
Oy 1274 -----AATGTCATGACATCTTATCCCTCCCAATCAGAGAGAGAGAGAG 1324
Db 369 HisArgGlnAsnGluIuProValPhe-----SerLysAspGlyArgLys 383
Oy 1325 CTCTGCTTTCCTCCGCGCCATGATGATCAAGACCGGCTTGTGCATTTGTACAAAGTCAC 1384
Db 384 PhePhePheIleArgValIlePro---GlnGlyGlyArgGlyLysPheYrHisIleThr 402
Oy 1385 GCGGTTTAAATCCACAGGCTACGATTGAGTGGACCGCTTCGCGCGCGGAGAGATGA 1444
Db 403 Val-----SerSerGlnProAsnSerAsnAspAsn--- 414
Oy 1445 TTTAACTGCCCATTAAGAGAGAGATTGCTGACACGCGGTGAATGGAGGTTTGGCG 1504
Db 415 -----IleGln-----SerIleThrSerGlyAspTrpAspVal----- 425
Oy 1505 AGGACGCGCTCCACAGATCTGGGTC---AATGAGAGACCAAGCTGCTACTTCCAGGCG 1561
Db 426 -----ThrLysIleLeuAlaTyrAspGluLysGlyAsnLysIleTyrPheLeuSer 442
Oy 1562 ACCAAGACACGCGCGCTGGAGCAGCAGCTCTACGTGTCATGAGAGGCGCGCGAG 1621
Db 443 ThrGlnAspLeuProArgArgArgGlnLeuTyrSerAlaAsnThrGluGly----- 459
Oy 1622 ATCGTACGCGCTCACACGCGCGCTTCTCCCA-AGCTGCTCCATGAGCCAG 1672
Db 460 -----AsnPheAsnArgGlnLysLeuSerGlyAspLeuValGlu 472
Oy 1673 AACTTGACATGTCCTCAGCCCTACAGCAGCGGTGAGCAGCGCGCTTCGTCGACGTC 1732
Db 473 AsnCysThrTyrPheSerIleAspSerPheSerHisSerMetAsp-----PhePheLeu 489
Oy 1733 TACAACTGACGCGCGCGCGCAG-----GACGACCGCGCTGCACAGCAG 1774
Db 490 LeuLysCysGlnGlyProGlyValIleProMetValThrValHisAsnThrAspLysLys 509
Oy 1775 CCGCGCTTGGGCTAGCATGATGAG---GCAGCAGCTGCGCGCGGATTAATGTTCC 1831
Db 510 LysMetPheAspLeuGluThrAsnGlnHisValLysLysAlaIleAsnAspArgGlnMet 529
Oy 1832 CCAGACATCTTCATTTCCACAGCGCG---TCGATGTCGCGCTCAGCGCATGATCTAC 1888
Db 530 ProLysValGluTyrArgAspIleGluIleAspAspLysAsnLeuProMetGlnIleLeu 549
Oy 1889 AAGCCCGACGCGCTTGACGCGAGGAGAAAGCAGCCCGCTCTTTGTATATGGAGCG 1948
Db 550 LysProAlaThrPheThrAspThrThrHisTyrProLeuLeuValValAspGlyThr 569
Oy 1949 CCGCAGGTGACGTGTGATTAACCTCTCAAGAGCATCAAGTACGCGCTCAACACA 2008
Db 570 ProGlySerGlnSerValAlaGluLysPheGlu---ValSerTrpGluThrVal---Met 587
Oy 2009 CTGGCTCCCTGGGCTACGCGCGGTGTGATGTAGCGGCGGCGCTCTGACGAGG 2068
Db 588 ValSerSerHisGlyAlaValValLysCysAspGlyArgGlyPheGlnGly 607
Oy 2069 CTTCGCTGCAAGGCGCGCGGAAACCAATGCGCAGGTGAGATCGAGGACGAGGT 2128
Db 608 ThrLysLeuLeuHisGluValAlaArgArgGluGlyLeuLeuGluGlyLysAspGlnMet 627
Oy 2129 GAGGCGCTGACGTCTGCGCGAGAGATAGCTCATGCACTGACCGCAGTTCAT 2188
Db 628 GluAlaValAlaArgThrMetLeu---LysGluGlnTyrIleAspArgThrArgValAlaVal 646
Oy 2189 CATGGCTGCTCAGGCGGCTCTCTGCTCATGCGGCTAATCCACAAG----- 2239
Db 647 PheGlyLysAspTyrGlyGlyTyrLeuSerThrTyrIleLeuProAlaLysGlyGluAsn 666

```

```

Oy 2240 ---CCCCAGTGTTCAGGTGGCATCGGGTGCCCGCGTACCGGTGTGATGGCTAC 2296
Db 667 GlnGlyGlnThrPheThrCysGlySerAlaLeuSerProIleThrAspPheLysLeuTyr 686
Oy 2297 GACACAGGTACAGTACGCGGTACATGAGCTGCGCTTCAGAACACAGCAGGCTATGAG 2356
Db 687 AlaSerAlaPheSerGluArgTyrLeuGlnHisGlyLeuAspAsnArgAlaTyrGln 706
Oy 2357 GCGGTTCCGTCGCGCTGACGCGGAGAGCTGCCCAATGAGCCCAAGCGCTTATC 2416
Db 707 MetThrLysValAlaHisArgValSerAlaLeu-----GluGlnGlnIlePheLeuIle 724
Oy 2417 CTTCACGCGCTTCCTGAGCAAAACGCTTTTCCACAAACCTTCTGCTGCCAA 2476
Db 725 IleHisProThrAlaAspGluLysIleHisPheGlnHisThrAlaGluLeuIleThrGln 744
Oy 2477 CTGATCCGACGAGGAAACCTTACAGCTCCAGATCTACCCCAAGACAGACAGTAT 2536
Db 745 LeuIleArgGlyLysAlaAsnTyrSerLeuGlnIleTyrProAspGluSerHisTyrPhe 764
Oy 2537 CGCTGCCCGAGTCGCGGCGAGCATATGAAGTACGCTGCTGCACTTCTACAGGAA 2593
Db 765 ThrSerSerLeuLysGlnHisLeuTyrArgSerIleIleAsnPheValGlu 783

RESULT 13
A:dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A41793
R:Wade, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthold, R.J.; Shimasaki, S.
Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A:Title: Differential expression of two distinct forms of mRNA encoding members of a
A:Reference number: A41793; MIMD:92108018; PMID:1729689
A:Accession: A41793
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-803 <MAD>
A:Cross-references: GB:M76429; NTD:9408719; PIDN:AAC41623.1; PID:9408720
A:Note: sequence extracted from NCBI backbone (NCBIP:75138)
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; glycoprotein
F:257/342/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 1,06e-18 Length: 803
Score: 442.50 Matches: 177
Percent Similarity: 38.99% Conservative: 108
Best Local Similarity: 24.21% Mismatches: 287
Query Match: 9.23% Indels: 159
DB: 2 Gaps: 28

US-09-976-674-4 (1-2617) x A41793 (1-803)
Oy 458 CTCTCTCTCTTCAGCGCCAGACAGCAGCTTTCACATGTCGCGAGCGGCAAGAGCG 517
Db 193 IlePheIlePheGlu-----AsnAsnIleTyrTyrCysAlaHisValGlyLysGlnAla 210
Oy 518 TTTCATGTGTCCCTATGAACCGCTGGAATCAAGACCCAGTGCAGGCGCGCGATG 577
Db 211 IleArgVal----- 213
Oy 578 GACCCCAAAATCTGCGCTGCGGACCTGCTTCTTCTTCATCAATAACAGCACTG 637
Db 213 ----- 213
Oy 638 TGAGTGCCAAATCAGACAGAGCGAGAGCGGCGCTGACCTTGTCCACCAAGGTTTA 697
Db 214 -----ValSerThrGlyLysGlnGlyValIle-----TyrAsnGlyLeu 226
Oy 698 TCCAAATGCTCGATGACCCCAAGTCTGCGGCTGTGGCCACCTTGTGTATACAGGAAG 757

```



```

RESULT 11
154331
dipeptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: I54331
R:Tokotani, N.; Doi, K.; Wentholt, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat
A:Reference number: I54331; MUID:93372805; PMID:8103397
A:Accession: I54331
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-865 <RES>
A:Cross-references: GB:M96859; NID:9306705; PIDN:AAA35760.1; PID:9306706
C:Superfamily: dipeptidyl-peptidase IV

Alignment Scores:
Pred. No.: 2,93e-19 Length: 865
Score: 452.00 Matches: 221
Percent Similarity: 39.96% Conservative: 149
Best Local Similarity: 23.87% Mismatches: 369
Query Match: 9.43% Indels: 187
DB: Gaps: 43

US-09-976-674-4 (1-2617) x I54331 (1-865)
QY 76 CCGCTCCAGGTGACAGAGCACTGTGGAGGCTCCGGAGCATCAT---CCAGGCAG 132
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 21 ProProGluAlaSerHisLeuGluGlyGlnGlyProGluGluAspGlyGlyAlaGly 40
QY 133 CCGCAAGTACTCGGCGCTATTGTCAACAAGCGCCCGACGACTTCCAGTTGTGCAGA 192
   ::  |||||  |||  |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 41 AlAlyProLeuGlyProArgAlaGlnAlaAlaProArg-----Glu 55
QY 193 GAGGATGAGTGGGCGCCACATCC-----CACCGCTTACTACTGGGAATGCC 243
   |||  |||  |||  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 56 ArgGly-GlyGlyGlyGlyAlaGlyAlaGlyArgProArgPheGlnTyrGlnIArgse 75
QY 244 ATATGCGAGCGGAGCAACACTCTCTCTACTGTAGATGCCAAGAAGTCCGGAAGA 303
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 75 rAspGlyAspGluGluAspGluLeuValGlySerAsnProGlnArgAsnTrpIysG 95
QY 304 GGCTCTGCTCTCTCTCTCTGTAAGAGATGTGATCATTTCCAGGCCAGCCACCA 363
   ::  |||  |||  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 95 yLeuAlaIleAlaLeuLeuValIleLeuValIleCysSerLeuIleValThrSerValI 115
QY 364 TGGGGTCTACTCTCGGAGAGAGAGCTGCTGAGAGAGCGGAACGCTGGGGTCTTGG 423
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 115 eLeuLeuThrProAlaGluAspAsnSerLeuSerGlnLysLysValThrValGluAs 135
QY 424 CATCACTCTCTACGACTTC-----CACAGCGAGAGTGGCTTCTCTTCCAGGCCAG 477
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 135 pLeuPheSerGluAspPheLysIleHisAspProGluAlaLysTrpIle-----Se 152
QY 478 CAACAGCTCTTCCACTGTCGAC-----502
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 152 rAspThrGluPheIleTyrArgGluGlnLysGlyThrValArgLeuTrpAsnValGluTh 172
QY 503 -----GGCGGCAAGAAC-----GGCTT 519
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 172 rAsnThrSerThrValLeuIleGluGlnLysLysIleGluSerLeuArgAlaIleArgTy 192
QY 520 CATGATGTCCTC-----ATGAACCGCTGGAAT 549
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 192 rGluIleSerProAspArgGlyIuThrAlaLeuPheSerTyrAsnValGluArgIle---Ty 211
QY 550 CAAGACCACTGCTCAGGCGCCCGGATGAGCCCAAAATGCTGCCCTGCCAGCTGCCCTT 609
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 211 rGlnHisSerTyrThrGlyThrGlyValLeuSerLysIleProHisGlyAspProGlnse 231
QY 610 CTTCCTCC-----TTGATCAATAAGAGGACCTGTGGGTGGCCAAATC---GAGACAG 660
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

Db 231 rLeuAspProProGluValSerAsnAlaLysLeuGlnIuThrAlaGlyTrpGlyProIysG 251
QY 661 CGAG-----GAGCGGCGGTGACCTTCTGC---CACCAAGATTATC 699
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 251 yGlnGlnLeuIlePheIlePheGluAsnIleTyrTyrCysAlaHisValGlyLysG 271
QY 700 C-----AATGTCCTGATGATGCCCAAGTCTGCG-----GGTGGCCACCTT 741
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 271 nAlaIleArgValIleSerThrGlyGlnGlyValIleTyrAsnGlyLeuSerAspTr 291
QY 742 CCGCATACAGGAAGAG---TTGACCGGCTTCACTGGGTACTGGTGGCCCAACAGCTC 798
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 291 pLeuTyrGlnGluGlnIleLeuLysThrHisIleAlaIleStrTrpSerProAspGlyTh 311
QY 799 CTGGGAAGTTACAGAGGCGCTCAAGACGCTGGCAATCTGTATGAGAGAGTGCATGAGTC 858
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 311 r-----ArgLeuAlaTyrAlaAlaIleAsnAspse 321
QY 859 CGAGGTGAGGTCAATTACAGTCCCTCT-----CTGCGCTAGACAGA 900
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 321 rArgValProIleMetGluLeuProThrTyrThrGlySerIleTyrProThrValLys-- 340
QY 901 AAGGAAGAGGAGCTGTATCGGTACCGCCAGAGCAGCAGCAAGATCCCAAGATTGCCCTT 960
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 341 -----ProTyrHisTyrProLysAlaGlySerGluAsnProSerIleSerIle 356
QY 961 GAACACTGTGATGTTCCAACTGACAGCCAGGCAAGATGCTCCAGCCAGAGAGA 1020
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 356 uHisValIle-----GlyLeuAsnGlyProThrHisAspLeuI 369
QY 1021 GCTGTGCAAGCCCTTCACTGCTGCTTCCCGAAGTGGAG-----TACATGCCAGGCG 1074
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 369 uMetMetProProAspAsp-----ProArgMetArgGluTyrTyrIleThrMetVa 386
QY 1075 CGGGTGACCCGGGATGAGCAATACGCGTGGGCCATGTTCTCCAGCCCGCCAGCAGTG 1134
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 386 LysTrpAlaThrSerThrLysValAla---ValThrPheAsnAlaArgAlaGlnsVa 405
QY 1135 GCTCCAGCTGCTGCTCTCTCCCGCGCCCTGTTCATCCGAGCAGCAAGATGAGAGCA 1194
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 405 lSerIleLeuThrLeuCysAspAlaThrThrGlyValCysThrLysLysHisGluAspG 425
QY 1195 GCGGTAGCCTCTGCCAGAGCTGTCCCAAGAAATGCCAGCCGTATGTGTACAGAGA 1254
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 425 u-----425
QY 1255 GGTCAACAAGCTGTGATC-----AATGTTCATGACATCTTATCCCTTCCCCA 1305
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 426 ---SerGluAlaThrPheHisArgGlnAsnGlnIuProValPhe-----439
QY 1306 ATCAGAGGAGAGAGAGAGAGCTGTTTCCGCGCCAAATGAAATGCAAGACGCGCTTGTG 1365
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 440 -SerLysAspGlyArgLysPhePheIleArgAlaIlePro---GlnGlyLysArgG 458
QY 1366 CCAATTGTCAAAAGTCAAGCGCGCTTTAAATCCAGAGGCTACGATTGAGATGAGCCCTT 1425
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 458 yLysPheTyrHisIleThrVal-----SerSerSerGlnProAs 471
QY 1426 CAGCCCGGGGAGAGATGAATTAAGTCCCATTAAGAGAAAGATGCTGTCGACAGCGG 1485
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 471 nSerSerAsnAspAsn-----IleGln-----SerIleThrSerG 483
QY 1486 TGAATGAGAGGATTTTGGCGAGCAGCGCTCCAAAGATCTGGGTC---AATGAGAGACCA 1542
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 483 yAspTrpAspVal-----ThrLysIleLeuAlaTyrAspGluLysGlyAs 498
QY 1543 GCTGTGTACTTCCAGAGCAGCAAGAGACGCGCTGAGACACCACTTACCTGTGTCAG 1602
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 498 nLysIleTyrPheLeuSerThrGluAspLeuProArgArgArgGlnLeuTyrSerAlaAs 518
QY 1603 CTATGAGCGCGCGGAGAGATGCTACGCGCTACACAGCGCGCGCTTCCCAT-----1654
   ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 518 nThrGlnGly-----AsnPheAsnArgGlnCyste 528

```

Db 142 sleuylsilevalasnasn-----gluarg1 152
QY 553 GACCCAGTGCATCAGGCGCCGATGAGACCCAAAATCGCCCTGCC-----GAGCC 603
Db 152 ethTyrAspIleelyleuvalglyslugluserValIleelnAlaPheIystrpansgl 172
QY 604 TGCCTTC-----TTCTCTTCATCAATTAACAGCGAGCTGTGGGTGGCCAAATCGAGAC 657
Db 172 yLysPheAsnAspPheValPheValIgluserAsnLysIleTyrIyrgInser-----se 190
QY 658 AGCGAGAGAGCGCGGTGACCTTCCTCCACCAAGTTTATCCATGCTCCGTGATGACC 717
Db 190 rProglInleuIleuIylerArValSerAsnIlyglIylnIhIsthrAlaSP----- 208
QY 718 CAAGTGCAGGTGTGCGACCTTCGTATACAGAGAGAG-----TTCCAGCGCTTCAGTGG 774
Db 209 -----GlyleuPheAspTrIleTyrIgluglulIlePheglIyArIyAspAl 225
QY 775 GTACTGTGTGCCCCACAGCTTCGGAGAGTTTCAGAGGCGCTCAAGAGCTGCGAAT 834
Db 225 aMetTrIyP-----SerThrLysglYAspGlN-----Le 235
QY 835 CCGTATAGAGAAAGTCATGATCCGAGGTGAGAGTCATTACGCTCCCTCT---CCTGC 891
Db 235 uAlaTyrAlaSerTyrAspAsnHisLeuThrLysAsnValSerleuLysThrTyrHisAr 255
QY 892 GCTAGAGAAAGAGAGAGCGAC---TGTATTCGTACCCGACGAGCAGCGAGCAAG---AA 945
Db 255 glenIuPrOtyrTrIleAspThrAsnPheHisTyrProLysThrPheAlaLysValLe 275
QY 946 TCCCAAGATTGCCTTAAGTGGCTGAGTTCCAGACTGACAGCCAGCGCAAGATCGTCTC 1005
Db 275 uProThrTyrThrleuSerIleTrpAsnLysIystrglu----- 288
QY 1006 GACCCAGAGAGAGAGCTG---GTCCAGCCCTTCAGCTCCGCTGTTCCCGAAGTGAGTA 1062
Db 289 ---GlnSerArgIleuAspValIleuLysAspSerLeu-----SerTyrHisIstY 305
QY 1063 CATCGCCAGGCGCGGTGG-----ACCCGGAGTGGCAATAGCGCTGGCG 1107
Db 305 rleuLeuAlaValIlystrPheugluIleasnIyThrgluIleuValSerValIyTrpH 325
QY 1108 CATGTTCCCTGAGCCGCGCCAGAGTGGTCCAGCTGCTCCTCCCGCGGCGCTGT 1167
Db 325 r-----AsnArgTyrGlnAsnIgluValAlaLeuThrIleGlyAspTrpAspThrAl 342
QY 1168 CATCCGAGACACAGACATAGAGACAGCGGCTAGCTTCGACAGAGCTGCCAGGAA 1227
Db 342 aileCysArgIleuIgluPheglu----- 349
QY 1228 TGTCCAGCCGTATGTGTGATCAGAGAGTCCACCAAGCTGTGATCAATGTATGCAAT 1287
Db 350 -----TyrLysTyrAlaSerLysArgTrpAl---ThrHisAspAs 362
QY 1288 CTTCATTCCTTCCTCCCAATCAGAGAGAGAGAGAGAGCTGTGCTTCGCGCGCAATGA 1347
Db 362 rPheHisSerIle-----ThrSerPheIuAspThrLeuPhePheleuLeuProHisAs 380
QY 1348 ATGCAAGACCGGCTTCGCATTTGTACAAGTCCAGCGCGTTTTAAATCCAGGCGTA 1407
Db 380 rLysArgAsp-----AsnAlaPheIgluIleValAlaSerleuArg----- 393
QY 1408 CGATTGGAGTACCCCTTCAGCCCCCGGGGAAGATGAATTAAGTCCGCCCAATTAGAGAGA 1467
Db 394 -----LeuSerHisIygluN-----LeuArgTyrTrp 402
QY 1468 GATTGCTGTGACAGCGGATATGAGGATTTTGGAGGACGAGCGCTCAAGATGTGGGT 1527
Db 402 oLysPheleuAsnleuIyIuTyrAspValIThrSerIleasnIy-----Il 418
QY 1528 CAATGAGAGACCAAGCTGTGACTTCAGGGCACCACAGAGACGCGCTGAGAGACCA 1587
Db 418 eaSnLysgluThrArgThrIlePhePheHisAlaIleAlaProLysProSerHisArgse 438

QY 1588 CCTTCAGTGTGATGATAGAGGCGCGGAGAGATGTACGGCTTCACCAAGCGCGGCTT 1647
Db 438 rleuPhe-----SerTyrSerleuAlaAspGlN-----SerArgAsnSerIatY 453
QY 1648 CTCCCATAGCTGTCTCATG-----AGCCAGAACTTCGACATGTT 1686
Db 453 rCysIleSerCysSerIleLysAsnCysThrTrpAlaIleIleValIleAspArgIle 473
QY 1687 CGTACGCCACTACAGCAGCGGTGAGACCGCGCGCTGGCTGCAC-----GTCTCAAGCT 1740
Db 473 LysThrAlaIleValSerCysLysglYProAlaIleProHisThrAlaIleValaSnle 493
QY 1741 GAGCGCGCCGACAGACAGCCCGCTGCACAAGCAG-----CCCGCGCTTCGT 1785
Db 493 uThrArgMetAspSerIyAspLysThrIgluHisAlaSnleuLeyTyrAspLysThrTy 513
QY 1786 GCGTACATGATGAGAGAGCAGCAGCTGCCCGCGATTAATGTTCTTCAGAGATCTTCA 1845
Db 513 rGlnAsnArgValIgluIleuIleValIleuPro-----ValIleIleLysgluThrIleY 531
QY 1846 TTTCACACCGCGCTCGATGTGCGGCTCTACGGCATGTATACAGCCCAAGCCTTGCA 1905
Db 531 sIleSerAspAspPheAspAlaLeuIleLysLeuSerIle-----ProLysAspIleTy 549
QY 1906 GCCAGGAGAGAGCAGC-----CCACAGCTCTTGTATAGAGGCGCCAGGT 1956
Db 549 rAsnAlaArgAspLysHisIleAlaIlePheleuIleValHisValIyTrIyglYProAsnAs 569
QY 1957 GCAAGCTGTGAATTAACCTCTTCAAA---GGCATCAAGTACTTGGCGGCTCAACACTGGC 2013
Db 569 rGlnAsnThrLysleuIleuIleThrIleIleIyglu-----GluValAlaI 585
QY 2014 CTCCCTGGGCTACGCGCTGTGTG---ATTAGCGAGGCGCTCTGTACGAGGCGCT 2070
Db 585 aSerAlaSerGlnAlaIleAlaIleuArgIleAspIyArgIySerIyglYArgIyTrgIyTr 605
QY 2071 TCGGTTCAAGGGGCGCTGAAAAACAAATGGCGAGCGATGGAGACAGGACAGCTGGA 2130
Db 605 rLysTyrArgSerIleAlaIleTyrIgluIleuIyThrValIgluValIuAspIleIly 625
QY 2131 GGGCTGTGACGTGTGCGCCAGAAATAT---GGTTCATCGACCTGAGCGGAGTGGCAT 2187
Db 625 sAlaIleIyValIleuArgLeuTyrArgHisIleuAspAlaArgIyAlaIaVa 645
QY 2188 CCAATGCTGTCTTACGCGGCGCTTCTCTGCTCATGCGGCTAATCCAAAGCC---CA 2244
Db 645 lPheglYTrpSerTyrIglYIlePheMetThrLeuSerMetValaSnIuIleProIgluI 665
QY 2245 GGTGTCAAGGTGGCCATCGCGGGTGGCGGCGTCCAGCTGTGGATGGCCAGACACAGC 2304
Db 665 rPhePheLysCysAlaValaSerValaIleProValIThrAsnPheAlaTyrTyrAspAlaTh 685
QY 2305 GTACACTGAGCGCTTACATG---GACGTCCCTGAGAACCAACAGCAGCGCTATGAGCGGG 2361
Db 685 rTyrThrGluArgTyrMetIyAspAlaPro-----LeuGl 697
QY 2362 TTCGTTGGCGCTCAGCTGAGAGAGTGGCCAAAT---GAGCCCAAGCGTTCATCTT 2418
Db 697 uSerTyrSerAspValIThrLysLysleuAspAsnPheLysSerThrArgleuLeuIeume 717
QY 2419 CCAAGGCTTCCTGAGCAAAAGCTGCACCTTTTCCACACAAACTTCTGCTCCCAACT 2478
Db 717 tHisglYleuLeuAspAsnValHisPheGlnAsnIleIleuIleAspIle 737
QY 2479 GATTCGAGAGGAGAAACCTTACAGGCTCGAGATCTACCCCAAGAGACAGAGATTTTCG 2538
Db 737 uGlnAsnArgIyValaAspPheAspIleuMetValIyTrpAsnIleHisSerleuSe 757
QY 2539 CTGCCCCGAGTGGGCGAGCATTGAAGTCAAGCTCAAGCTCTGCACTTCTTACAGGAA 2593
Db 757 rSerArgThrSer-----HisValIyglYsMetThrHisPheleuArgIln 773

QY	1135	GCTCAGCTGCTCTCTCTCCCGGCGGCGCTTCACATCCGCGACACAGGAATGAGSACA	1194
Db	331	uvalalaleuthrllecyasprtrpsarphralaleicysargleuglupheul----	369
QY	1195	GCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAATGTCCAGCGCTATGGTGTCAGSAGA	1254
Db	370	-----TyrLeuTy 372	
QY	1235	GCTACCAACAGCTCGGATTCATGTTATGACATCTTCTATCCCTTCCCAATCAGAGG	1314
Db	372	ralaserlysartrpyval---Thrlsasprphenlsertile-----ThSerph 389	
QY	1315	AGAGACAGACGTCTGCTTCTCCGCGCAATGACAAATCAGACCGGCTTCTGCATTTGTA	1374
Db	389	egluasprhlreupheheuleuLeuprhnsarplysargap-----AsnAlarph 406	
QY	1375	CAAACTCAGCGCGTTTAAATCCAGAGGCTACGATTGAGTGAAGCCCTTCAGCCCGG	1434
Db	406	eglnlgnlvalalaserleuarf-----LeuSerhls61 417	
QY	1435	GGAAGATGAATTTAAGTCCCATTAAGAAAGATGTGCTGACACAGCGGATGAGGA	1494
Db	417	ylGln-----LeuarghlrProlysrPhleuAsnleuglyglutyrAs 431	
QY	1495	GCTTTGGCGAGCGAGCGTCCAAAGATCTGGTCAATGAGAGACCAAGCTGTACTT	1554
Db	431	pvalrhtserlserlensnly-----lIeasnlysgluthrPhagThrlePherph 447	
QY	1555	CGAGGCGACCAAGACACAGCGCGTGGAGACACACACCTACGTACGTACATGAGCGCG	1614
Db	447	ehlsalalalalrProlysrProserhlsarSerleuPhe-----SerTyrSerleual 465	
QY	1615	CGCGAGATCTTACGCGCTTCACACAGCGCGGCTTCTCCATAGCTGTCTCATG----	1666
Db	465	asaprlu-----SerlragsnserlArlyrcyslIeasercyslerlIeasrcy 482	
QY	1667	-----AGCCAGAACTTCGACATGTTCTGACGCCACTACAGACGCGTGAGCAC	1713
Db	482	sthTrpAlaGlnlAglInmetaspsrplnmetlstrhlralalevalsercyslysgl 502	
QY	1714	GCCGCGCTGCGCTGAC-----GTCACAAAGCTGACGCGCGCCGACGAGACCGCTTCA	1767
Db	502	yrProhlaalarohlsrhlralalevalasnlleuhtlrArghmetasbersarplyslyth 522	
QY	1768	CAAGCAG-----CCCGCTTCGGGCTAGCATGATGAGAGCAGCCAGCTG 1812	
Db	522	rglInhlslaslnleuLeuLeuLyrsprlthryrglnasnarAgvalgluglualaglyle 542	
QY	1813	CCCCCGGATTAATGTCTCTCCAGAGATCTCCATTTCCACAGCGCGCTGGAGTGGCGCT	1872
Db	542	uPro-----VallellelleysgluthrlllelyslleasrapsarphenaspalaLeuul 560	
QY	1873	CTAGCGCATGTACTTCAACAGCCCAACGCGCTTGACGCGAGGAAGAACAC-----CC	1923
Db	560	elysleuserlile-----ProlysarplyletryAsnarAgasrplyshlsGlnAlailePr 578	
QY	1924	CACGCTCTCTTTTATATATGAGAGCCCAAGGTGACGTGGTGAATTACTCTTAA--	1961
Db	578	oleuLeuAlahlsValtyrlygllylProbasnspglnasnthrlysglualalThglnll 598	
QY	1982	GGCATCAAGTACTTGGCGGTCAACACACTGAGCGCTCCGCGGTACGCGGTGGTG--	2038
Db	598	eglylIlelu-----GluvalvalAlaserialaserGlnAlaAlaIleleuar 614	
QY	2039	ATTGACGGCAGGGGCTCTGTACAGCAGAGGCTTCGTTTGAAGGGCCCTGAAAAACA	2097
Db	614	gllaspolyArgetlyserlyelylArgetlytrpLysTyAraserAlailetyrlygl 634	
QY	2098	AATGGCGGACGTGAGATTCAGAGGACACAGGTGAGGGCGCTGACGTTGTGGCGGACAACTA	2157
Db	634	nlneuglyhrValolGluvalGlnasprlnllelysalalellysalvalleuAlagLeuTy 654	

QY 2158 T--GGCTTACACCGCTAGCCGAGTGTGGCATTGCATTCACCTGGCTCTACGGGGGGCTTCT 2214
Db 654 TATGHTISLeuLeuaspIlaarIgarIyAlaIaVal PheIlytPserTYrGIyGIyPheMe 674
QY 2215 CTCGCTCATCGGGGTAAATCCACAAGAGCCC--CAGGTGTTCAAGGTGGCCATGGCGGGGTGC 2271
Db 674 tThrLeuSerMetValasnGlnIalProGlnInPhePheIyCysAlaValSerValAl 694
QY 2272 CCGCGTCAACCGCTTGATGGCCTTACGACACAGGGTACACTGAGCGCTATACG--GAGCT 2328
Db 694 aProValThrsInpheaIaTYrIyAspAlaThrTYrInGlnIaTYrIyMetGIyAspAl 714
QY 2329 CCGTGAAGAACAACAGACGAGGATAGAGCGGGGTGGCGGCGCCGTCACAGCTGGAGAACCT 2388
Db 714 aPro-----LeuGlnSerTYrSerAspValThIyIyIyIy 726
QY 2389 GCCCAAT--GAGGCCAACCGCTTGATTCTCCACAGCGCTTCGTGGAGAAACGTGCA 2445
Db 726 uAspInpheaIaTYrIyAspAlaThrIyLeuIleuMetHisGlyLeuIleuAspAspAsnValHI 746
QY 2446 CTTTTCACACAAACTCTCTCGTCTCCCACTGATCCGACGAGGAAACCTTACCAGCT 2505
Db 746 sPheaInasnSerIaIleLeuIleIleAspGlnIleuGlnasnIarGlyIaIaAspPheaSple 766
QY 2506 ICAGATCAACCCCAACAGACAGACAGATATCTCTCCCGAGTGGCGGAGCACTATGA 2565
Db 766 uMetValIyProAsnGlnIalHisSerIleuSerSerIarGlnIser-----HisValIa 784
QY 2566 AGTCACGTTGCTGCACTTTCTACAGAA 2593
Db 784 IGIyIyMetThrHisPheLeuArgIn 793

RESULT 10
 T25173
 hypothetical protein T23F1.7a - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T25173
 R:Wilkinson, J.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19990
 A:Accession: T25173
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-779 <MWt>
 A:Cross-references: EMBL:Z81129; PIDN:CAB03411.1; GSFDB:GN00023; CESP:T23F1.7a
 A:Experimental source: clone T23F1
 C:Genetics:
 A:Gene: CESP:T23F1.7a
 A:Map position: 5
 A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1
 C:Superfamily: dipeptidyl-peptidase IV

[illegible]

US-09-976-674-4 (1-2617) x T25173 (1-779)

[illegible]

```

QY 1760 CCCCTGACAGACGCCCGCTCTGGGCTAGCATGATGAGGAGCCAGCTGC----- 1813
      |||||
      |||||
      |||||
Db 482 -----HisylsGluLeuArg-----ValLeuGluAspAsnSerAlaLeuAsp 495
QY 1814 -----CCCCGGATTATTCCTCCAGAGATCTTCATTTCCACAGCCGCTGGAT 1864
      |||||
      |||||
      |||||
Db 496 ArgMetLeuGlnAspValGlnMetProSerLysLeuAspPheIleValLeuAsnGlu 515
QY 1865 GTGGGCGCTC---TAGCGCATGATCTACAGCCCGCTTCGAGCCAGGAGAACAGAC 1921
      |||||
      |||||
      |||||
Db 516 ThrArgPheThrPyrGlnMetIleLeuProPheIle---PheAspLysSerLysLysTyr 534
QY 1922 CCCACCGCTCTTTGTATATGAGGCCCGCCAGGTGACGTGTAATTAACCTTCGAA 1981
      |||||
      |||||
      |||||
Db 535 ProbenLeuAspValTyrAlaGlyProCysSerGlnLysAlaAspAlaSerPhe--- 553
QY 1982 GGCATCAAGTACTGGCGCTCAC-----ACACTGGCTCTCCCTGGGCTACGCCGTG 2022
      |||||
      |||||
      |||||
Db 554 -----ArgLeuAsnTrpAlaThrTyrLeuAlaSerThrGluAsnIleIle 568
QY 2033 GTTGTG---ATTGACGGCAGGGGCTCTGTCAGCGAGGCTTCGAAAGGGCCCTG 2089
      |||||
      |||||
      |||||
Db 569 ValAlaSerPheAspGlyArgGlySerGlyTyrGlnGlnAspLysIleMetHisAlaIle 588
QY 2090 AAAAACCAATGGCGCAGGTGAGATCGAGACCGACCGGTGAGGGCGCTG---CACTTCGTG 2146
      |||||
      |||||
      |||||
Db 589 AsnArgArgLeuGlnGlyThrLeuGlnValGlnAspGlnIleGlnAlaAlaArgGlnPheVal 608
QY 2147 GCCGAGAGATAGTGGCTTCATCGACCTGACCGGATGGCATCCATGGCTGCTCTACGGG 2206
      |||||
      |||||
      |||||
Db 609 -----LysMetGlyPheValAspSerLysArgValAlaIleTrpGlyTrpSerTyrGly 626
QY 2207 GGCCTCCCTCCGCTGAGGGGCTATATCACAAGCCCGAGGCTTCAGAGTGGCCATGCGC 2266
      |||||
      |||||
      |||||
Db 627 GLTyrValThrSerMetValLeuGlySerGlySerGlyValPheLysCysGlyIleAla 646
QY 2267 GGTGCCCCGCTCAACCTGTGATGGCTTACGACACAGAGGTACACTGACGCTACATG--- 2323
      |||||
      |||||
      |||||
Db 647 ValAlaProValSerArgTrpGlyTyrLysPheSerValTyrThrGlnArgTyrMetGly 666
QY 2324 ---GACCTCCCTGAGAACACACAGCGGTATGAGCGGGTCCGTCGGCTGACAGTGTG 2380
      |||||
      |||||
      |||||
Db 667 LeuProIleProGlnAspAsnLeuAsnLysTyrArgAsnSerThrValMetSerArgAla 686
QY 2381 GAGAAAGTGGCCCAATGAGCCCAACCGCTGCTTATCTCCACAGGCTTCGAGGAAAC 2440
      |||||
      |||||
      |||||
Db 687 GlnHisPhe-----LysGlnValGlyTyrLeuLeuIleHisGlyThrAlaAspAsn 704
QY 2441 GTGCATTTTCCACACAAACTCTCTGCTCCCACTGATCCGAGGAGAAACCTTAC 2500
      |||||
      |||||
      |||||
Db 705 ValHisPheGlnGlnSerAlaGlnIleSerLysValLeuValAspAlaGlyValAspPhe 724
QY 2501 CACTCCAGATCTATCCCAAGAGAGACAGATATGGCTGCCCGGAGTGGCGGAGCAC 2560
      |||||
      |||||
      |||||
Db 725 GlnAlaMetTrpTyrThrAspGlnAspHisGlyIleAlaSerSerThrAlaHisGlnHis 744
QY 2561 TATGAAGTCACTGCTGCTCACTTTCTACAGAA 2593
      |||||
      |||||
      |||||
Db 745 IleTyrSerHisMetSerHisPheLeuGlnGln 755

```

```

A:Cross-references: EMBL:Z81129; PIDN: CAB03412.1; GSPDB: GN00023; CESP: T23F1.7b
A:Experimental source: clone T23F1
C:Genetics:
A:Gene: CESP: T23F1.7b
A:Map position: 5
A:Insertions: 13/3; 52/1; 111/3; 197/3; 366/3; 524/1; 557/1; 648/2; 703/1; 743/1
C:Superfamily: dipeptidyl-peptidase IV

Alignment Scores:
Pred. No.: 5 38e-21 Length: 799
Score: 479.50 Matches: 207
Percent Similarity: 42.03% Conservative: 125
Best Local Similarity: 26.20% Mismatches: 302
Query Match: 10.00% Indels: 156
Gaps: 44

US-09-976-674-4 (1-2617) x T25174 (1-799)
QY 385 GAGACCTGCTGAGGAGCGGAA---ACGCTGGGGGTCTGCGCATCCTCTACGACTT 441
      |||||
      |||||
      |||||
Db 107 GlyArgGlnAsnGlyAlaGluIleThrProSerAlaAspArgLysTyrPheAlaMetMet 126
QY 442 CCACAGCGAGAGTGGCTTCTCTCTCCAGC-----CAGCAACAGCTCTTCCA 492
      |||||
      |||||
      |||||
Db 127 AspHis-----AlaProAsnProGlyMetAsnProGln-AsnGlnThrPheI 142
QY 493 CTGTGCGGAGCGCGGCAAGACGGCTTCATGCTGCTCCATGAAACCGGTGAGAAATCAA 552
      |||||
      |||||
      |||||
Db 142 SLeuLysIleValAlaAsnAsnGln---ArgLeuAsnProLeuLeuProPheGluValG 161
QY 553 G-----ACCCAGTGTCTACGAGCCCGGATGCA 579
      |||||
      |||||
      |||||
Db 161 uGlnLeuPheArgGlnLeuSerAspSerArgIleThrTyrAspIleGlyLeuArgLysG 181
QY 580 CCCCATAATGCTGCCCTGCC-----GACCTGCTTC-----TTCTCTTCATCAA 624
      |||||
      |||||
      |||||
Db 181 uGlnSerValIleGlnAlaPheLysTrpAsnGlyLysPheAsnAspPheValPheValG 201
QY 625 TAACAGCGACCTGGGTGGCGCAACATCGAACAGAGGAGGAGCGGGCGGTGACCTCTG 684
      |||||
      |||||
      |||||
Db 201 uSerAsnIleLeuTyrGlnSer-----SerProGlnGlnGlyLeuLeuThrArgVal 219
QY 685 CCACACAGTATTATCAATGCTCTGTGATGACCCCACTGCTGCGGTGCGCACTTCGT 744
      |||||
      |||||
      |||||
Db 219 LserAsnGlyGlyGlnHisThrValAsp-----GlyLeuPheAspTrpI 234
QY 745 CATACAGAGAGAG---TTCACGCGCTTCACTGGTACTGGTGGTGGCCACAGCTCTGTG 801
      |||||
      |||||
      |||||
Db 234 eTyrGlnGlnGlnIlePheGlyArgLysAspAlaMetTrpTrp-----SerTh 250
QY 802 GGAAGGTTCAGAGGGGCTCAAGACGGTGCAGATCTGTATGAGAGAGATGATGATGCGCA 861
      |||||
      |||||
      |||||
Db 250 rLysGlyAspGln-----LeuAlaTyrAlaSerTyrTrpAsnHisIle 264
QY 862 GGTGAGGATCATTCACGTCCCTCT---CCTGGCTAGAGAAAGAAAGAGAGAC---TC 915
      |||||
      |||||
      |||||
Db 264 uThrLysAsnValSerLeuLysTyrThrArgLeuGlnProTyrProIleAspThrAs 284
QY 916 GTATCGGTACCCACAGGACAGGACAGAG---ATCCCAATTCCTTGAACCTGCTGA 972
      |||||
      |||||
      |||||
Db 284 nPheHisTyrProLysThrPheAlaLysValLeuProThrTyrThrLeuSerIleTrpAs 304
QY 973 GTTCCAGACTGACAGCCAGGAGCAAGATCGTGTGACCCAGAGAGAGAGGTG---GTGCA 1029
      |||||
      |||||
      |||||
Db 304 nLysLysThrGln-----GlnSerArgGlnLeuAspValG 316
QY 1030 GCCCTTACGCTGCTGTTCGCCAAGGTGAGTACATCGCCAGGAGCGCGGTGG----- 1081
      |||||
      |||||
      |||||
Db 316 nLeuLysAspSerLeu-----SerTyrHisTyrLeuLeuAlaValLysTrpLeuGluI 334
QY 1082 -----ACCCGGAGAGCAAAATACGCTGGCGCATGTTCTGACCGCCCGACGACGTG 1134
      |||||
      |||||
      |||||
Db 334 eAsnGlyThrGlnGlnLeuValSerValTrpThr-----AsnArgTyrGlnAsnG 351

```


QY 2426 TTCCTGGAGAAAACGTGACCTTTTTCACACAAACTCTCTCTCCCAACTGATCCGA 2485
||||:|||||
Db 706 ThrAlaAspAsnValHisPheGlnInSerAlaGlnHisLeuLysAlaLeuValAsp 725
QY 2486 GCAGGAAACCTTACCACTCCGATCTACCCACGAGACGATCTTCCTCCCTCCCC 2545
|||
Db 726 ValGlyValAspPheGlnAlaMetTrpTyrThrAspGlnAspHisGlyLeuAlaSerSer 745
QY 2546 GAGTCGGGGAGACATGAGTACGACCTGCTGCTCTCTCTACAGGAA 2593
Db 746 ThrAlaHisGlnHisIleTyrThrHisMetSerHisPheIleuSerGln 761
RESULT 7
A39914
dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
M:Alternate names: GP110; membrane glycoprotein 110K; OX-61
N:Contents: dipeptidyl-peptidase IV, soluble form
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #text, change 10-Sep-1999
C:Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781
R:Hong, W.; Doyle, D.
Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987
A:Title: cDNA cloning for a bile canalliculus domain-specific membrane glycoprotein of rat
A:Reference number: A39914; MUID:88068516; PMID:3479775
A:Accession: A39914
A:Molecule type: mRNA
A:Residues: 1-792 <HON>
A:Cross-References: GB:J02997; NID:9204463; PIDN:AAA1272.1; PID:9204464
R:Ogata, S.; Misumi, Y.; Ikehara, Y.
J. Biol. Chem. 264, 3596-3601, 1989
A:Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA an
A:Reference number: A33315; MUID:89123496; PMID:2563382
A:Accession: A33315
A:Molecule type: mRNA
A:Residues: 1-37, 'A', 39-182, 'I', 184-331, 'T', 333-351, 'C', 353-393, 'V', 395-561, 'L', 563-623,
A:Cross-References: GB:J04591; NID:9203973; PIDN:AAA1096.1; PID:9203974
A:Note: The authors translated the codon GCG for residue 38 as Arg, ACC for residue 332
A:Accession: B33315
A:Molecule type: protein
A:Residues: 1-20;35-54;427-443;505-509;511-520;530-538;593-600;602-608;618-627 <HO2>
R:McCaughan, G.W.; Mickson, J.E.; Creswick, P.F.; Gorrell, M.D.
Hepatology 11, 534-544, 1990
A:Title: Identification of the bile canallicular cell surface molecule GP110 as the ectop
peptide.
A:Reference number: A60730; MUID:90228896; PMID:1970322
A:Accession: A60730
A:Molecule type: protein
A:Residues: 28-47, 'X', 50-53, 55-58 <MCC>
R:Ogata, S.; Misumi, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.
Biochemistry 31, 2582-2587, 1992
A:Title: Identification of the active site residues in dipeptidyl peptidase IV by affini
A:Reference number: A42203; MUID:92190188; PMID:1347701
A:Accession: A42203
A:Molecule type: protein
A:Residues: 'R', 625-630, 'X', 632-648 <OG2>
R:Iwaki-Egawa, S.; Watanabe, Y.; Fujimoto, Y.
Biol. Chem. Hoppe-Seyler 374, 973-975, 1993
A:Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl p
A:Reference number: S38949; MUID:94128239; PMID:7905271
A:Accession: S38949
A:Status: preliminary
A:Molecule type: protein
A:Residues: 281-302 <IWA>
R:Hong, W.; Doyle, D.
J. Biol. Chem. 263, 16892-16898, 1988
A:Title: Membrane orientation of rat gp110 as studied by in vitro translation.
A:Reference number: A31781; MUID:89034185; PMID:3182821
A:Accession: A31781
A:Molecule type: mRNA
A:Residues: 1-40 <HO3>
C:Comment: This protein is localized to the bile canalliculus, which is the apical domain
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptidase hydrolase; glycoprotein; homodimer; liver; serine protease

F:1-792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATM
F:1-28/Domain: signal sequence #link MATS #status experimental <SIG>
F:1-6/Domain: intracellular #status predicted <INT>
F:7-28/Domain: transmembrane #status predicted <TMN>
F:29-792/Domain: extracellular #status predicted <EXT>
F:29-94/Domain: propeptide #link MATS #status experimental <PRO>
F:35-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MAT>
F:83,90,148,217,227,319,521,686/Binding site: carbohydrate (Asn) (covalent) #status p
F:631/Active site: Ser #status experimental
F:709,741/Active site: Asp, His #status predicted

Alignment Scores:

Pred. No.:	3,19e-21	Length:	792
Score:	484.50	Matches:	205
Percent Similarity:	37.74%	Conservative:	123
Best Local Similarity:	23.59%	Mismatches:	334
Query Match:	10.10%	Indels:	207
DB:	1	Gaps:	36

US-09-976-674-4 (1-2617) x A39914 (1-792)

QY 197 GATGACTCTGGGCCCCCAGCTCCACCGCTCTTAC-----TACCTGGGAATGCCA 244
|||||
Db 31 AspGluAlaAlaAspSerArgTrpTyrThrLeuAlaAspTyrLeuLysAsnThr 50
QY 245 TATGGCAGCGAGAGAAAGTCCCTCTTAC---TCTGAGATTCCAAGAGTCCGGGANA 301
|||
Db 51 PheArgValLysSerTyrSerLeuArgTrpValSerAspSerGluTyrLeuTyrGln 70
QY 302 GAGCTCTGCTGCTCTCTCTCTCTGGAAGCAGATCTGATCATTTCCAGCCACGCCCCAC 361
|||
Db 71 GluAsnAsnIleLeuLeu-----PheAsnAla-----Glu 80
QY 362 CATGGGGTCTACTCTCTGGGAGAGAGACCTCTGAGGAGCGGAAGCGCTGGGGCTCTC 421
|||||
Db 81 HisGly-----AsnSerSerIlePheLeuGlnAsnSerThrPheGluIlePhe 96
QY 422 GGCATACCTCTACGACATCTCCAC---ACGAGAGTGGCTCTCTCTCTCCAGGCCACG 478
|||
Db 97 GlyAspSerIleSerAspTyrSerValSerProAspArgLeuPheValLeuGluTyr 116
QY 479 AACAGCCTCTTCACGTGCGCGAGCGCGCAGACAGCGCTTCATGGTGTCCCTATGANA 538
|||
Db 117 AsnTyrValLysGlnTrpArgHisSerTyrThrAlaSerTyrSerIleTyrAspLeuAsn 136
QY 539 CCGCTGGAATCAAGACCAGCTGCTCAGGCGCCCGGATGAGACCCCAATCTGCCCTGCC 598
|||
Db 137 LysArgGlnLeuIleThrGlnGluLysIleProAsnAsnThrGlnTrpIleThrTrpSer 156
QY 599 GACCTGACCTTC---TTCCTTCATCAATTAACAGGACCTGTGGGCGCAACATCGAG 655
|||
Db 157 GlnGlnGlyHisLysLeuAlaTyrValTrpLysAsnAspIleTyrValLysIleGluPro 176
QY 656 ACAGGGAGAGGCGGCGCTGACCTTCTGCCACCAAGTTTATCAATGCTGTGATGAC 715
|||
Db 177 HisLeuProSerHisArgTrpThr-----SerThrGlyGlnLysAlaValIlePheAsn 194
QY 716 CCAAGTCTGGGGTGTGGCCACCTTCATACAGGAAGAG---TTCGACCGCTTCAC 772
|||
Db 195 -----GlyIleAsnAspTrpValTyrGlnGluLeuIlePheGlyAlaTyrSer 210
QY 773 GGCTACTGGTGGCGCCACAGCGCTCTGGAGAGGTTCAGAGGGCTCAAGACGTGCGA 832
|||||
Db 211 AlaLeuTrpTrpSerProAsnGlnTyrPhe----- 220
QY 833 ATCTCTATGAGGAAGTGCATGATGCCAGGTGAGCTCATTCAC-----GTCCCC 883
|||
Db 221 LeuAlaTyrAlaGlnPheAsnAspTrpGlyValProLeuIleGluTyrSerPheTyrSer 240
QY 884 TCTCCTGGCTAGAGAAAGAGAGAGAGAGCTGTATCGGATACCCCGAGGACGACCAAG 943
|||
Db 241 AspGluSerLeuGlnTyrProLysTrpValTrpIleProTyrProLysAlaGlyAlaVal 260

Db 96 AspGluPheGlyHisSerIleAsnAspTyrSerIleSerProAspGlyGlnPheIleLeu 115
QY 470 CAGCGCAACAACAGCTCTTCCACTGTGCGACGGCGCAAGACGGCTTCATGCTGTC 529
Db 116 LeuGluTyrAsnTyrValLysGlnTPrArgHisSerTyrThrAlaSerTyrAspIleTyr 135
QY 530 CCAATGAACCGGTGAAGATCAAGACCAGTGTCCAGGCCCCGAGATGAGCCCAAAATC 589
Db 136 AspleuAsnTyrArgGlnLeuIleThrGlnGluArgIleProAsnAsnThrGlnTPrVal 155
QY 590 TGCCCTCCGACCCCTGGCTTC---TTCCTCTCAATCAATACAGACGACCTGGGTGGC 646
Db 156 ThrTPrSerProValGlyHisLysLeuAlaTyrValTPrAsnAsnAspIleTyrValLys 175
QY 647 AACATCAGACAGACGAGGAGGCGGCTGACCTTCTGCCACCAAGGTTTATCCATGTC 706
Db 176 IleGluProAsnLeuProSerTyrArgIleThrTPr-----ThrGlyLysGlnAspIle 193
QY 707 CTGGATGACCCCAAGTCTGGGGGTGGCCACCTTCGTACATACAGAAAG---TTCGAC 763
Db 194 IleTyrAsn-----GlyIleThrAspTPrValTyrGlnGlnGluValPheSer 209
QY 764 CGCTTCACCTGGGTACTGTGTGGTCCCAACAGCTCTCGGAGAGTTCAGAGGGCTCAAG 823
Db 210 AlaTyrSerIleLeuTPrTPrSerProAsnGlyThrPhe----- 222
QY 824 ACGGTGCAATCCTGTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 877
Db 223 -----LeuAlaTyrAlaGlnPheAsnAspThrGluValIProleuIleGluTyrSer 239
QY 878 ---GTCCCTCTCTCGCTAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 934
Db 240 PheTyrSerAspGlnSerLeuGlnTyrProLysThrValArgValProTyrProLysAla 259
QY 935 GCGAGCAAGATCCCAAGATTTGCTTGAACCTGCTGATTCAGACTGACAGCCAGGCG 994
Db 260 GlyAlaValAsnProThrVal-----LysPhePheValValAsnThrAspSerLeuSer 277
QY 995 AAGATCTCTCGACCCAGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1054
Db 278 SerValThrAsnAlaThrSerIleGlnIleThrAlaProAlaSerMetLeuIleGlyAsp 297
QY 1055 GTGGAGTACATCGCCAGGCGGGGTGACCCGGGATGGAATACAGGCTGGCGCATGTC 1114
Db 298 ---HisTyrLeuLysAspValThrTPrAlaThrGlnGlnGluArgIleSer----- 312
QY 1115 CTGGACCGGGCCCAAGAGTGGCTCCAGCTGCTCTCTCCCTCCCGCGGCTTCATCCCG 1174
Db 313 -----LeuGlnTPrLeu----- 316
QY 1175 AGCACAAGATGAGAGAGAGCGGTAGCCTTCGACAGACTGTCGCCAGATGTCAG 1234
Db 317 -----ArgArgIleGln 320
QY 1235 CCGTATGTGTCG-----TACGAGAGGTACACCAACGTCGTGATCAATGTT 1279
Db 321 AsnTyrSerValMetAspIleLysAspTyrAspLeuSerGlyArgTPr----- 337
QY 1280 CARGACATCTTCATCCCTTCCCAATCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1339
Db 337 ----- 337
QY 1340 GCCAATGATGACAAGACCGGCTTCTGCAATTTGTACAAAGTACACCGCGTTTAAATCC 1399
Db 338 -----AsnLysLeuValAlaArgGlnHisIleGluMetSerThrTPr----- 351
QY 1400 CAGGGCTACATTTGAGTGAAGCCCTTCAGCCCGGGGAGAGATGA----- 1444
Db 352 -----GlyTPrValGlyArgPheArgProSerGlnProHisPheThrLeuAspGly 368
QY 1445 -----TTTAAG 1450
Db 369 AsnSerPheTyrLysIleIleSerAsnGlnGluGlnTyrTArgHisIleLysTyrPheGln 388

QY 1451 TGCCCAATTAAAGAGAGATTCCTGTGACACCGGTGATGAGAGATTGGCGAGGAC 1510
Db 389 IleAspLysLysAspCysThrPheIleThrLysGlyThrTPrGlnValIleGlyIleGlu 408
QY 1511 GCGTCCAGATCTGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGCGCAAGAGAC 1570
Db 409 AlaLeu----- 1570
QY 1571 ACGCGGTGAGACCAACACCTCTAGTGTCATGATGAGCGGGCGGC----- 1618
Db 414 -----TyrLeuTyrTyrIleSerAsnGlnTyrLysGlyMetProGlyGly 428
QY 1619 ---GAGATCGTACCGCTTCACACCGCGGCTTCTCCCATAGCTGTCATGAGCGCAAAAC 1675
Db 429 ArgAsnLeuTyrLysIleGlnLeuIleAspTyrThrLysValThrCysLeuSerCysGlu 448
QY 1676 TTGCAATGTTCTGTCACCATACAGACAGCTGAGACCGCGCGCTGTCACAGCTTAC 1735
Db 449 LeuAsnProGluArgCysGlnTyrTyrSerValSerPheSerLysGlnAlaLysTyrTyr 468
QY 1736 AAGCTG-----ACGGCGCC----- 1750
Db 469 GlnLeuArgCysSerGlyProGlyLeuProLeuTyrThrLeuHisSerSerValAsnAsp 488
QY 1751 -----GACGACAGCCCGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1789
Db 489 LysGlyLeuArgValLeuGlnAspAsnSerAlaLeuAspLys----- 502
QY 1790 ACGATGATGAG 1849
Db 503 ---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe 516
QY 1850 CACACGGCTCGGATGTGGCGTC---TACGGCATGATCTACAGCCCGACGCTTGCAG 1906
Db 517 IleIleLeuAsnGlnTyrLysPheTyrTyrGlnMetIleLeuProProHis---PheAsp 535
QY 1907 CCAGGAGAAAGACACCCACCGCTCTTTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1966
Db 536 LysSerLysLysTyrProLeuLeuLeuAspValTyrAlaGlyProCysSerGlnLysAla 555
QY 1967 AATATCTCTTCAAAAGCATCAAGTACTTGGCGCTCAAC-----ACATGGCGCTCC 2017
Db 556 AspThrValPhe-----ArgLeuAsnTPrAlaThrTyrLeuAlaSer 569
QY 2018 CTGGGCTACCGCGCTGTGTC---ATTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2074
Db 570 ThrGlnAsnIleValAlaSerPheAspLysArgLysGlyTyrGlnGlyAspLys 589
QY 2075 TTGCAAGGGCGCTGAAACCAATGAGCCAGGTGAGATGAGAGAGAGAGAGAGAGAGAG 2134
Db 590 IleMetHisAlaIleAsnArgArgLeuGlyThrPheGlnValGlnAspGlnIleGlnAla 609
QY 2135 CTG---CAGTTCGTGGCGGAGAGATGAGCTTATGCACTGAGCCGATGTCATCCAT 2191
Db 610 AlaArgIlePhe-----SerLysMetGlyPheValAspAsnLysArgIleAlaIleTPr 627
QY 2192 GCGTGTCTACAGGGGGCTCTCGCTCATAGGGGTATTCACAGCCCGAGGTGTC 2251
Db 628 GlyTPrSerTyrGlyGlyTyrValThrSerMetValLeuGlySerGlySerGlyValPhe 647
QY 2252 AAGGTGCGCATCGCGGGGTGCGCGGATGACAGCTGTGAGTGGCTTACAGACAGGATAC 2311
Db 648 LysCysGlyIleAlaValAlaProValSerArgTPrGlyTyrLysPheSerValTyrThr 667
QY 2312 GAGCGTACATGAGAGTC-----CTGAGAACACACAGACAGGCTATGAGCGGGTTC 2365
Db 668 GlnTyrGlyMetGlyLeuLeuProThrProGlnAspAsnLysAsnIleTyrArgAsnSerThr 687
QY 2366 GTGGCGCTGACAGTGGAGAGAGCTGCCAATGAGGCCCAAGCGCTTGTATTCCTCAGGCG 2425
Db 688 ValMetSerArgAlaGlnAsnPhe-----LysGlnValGlnTyrLeuLeuIleHisGly 705

Db 645 -----LysGlyLysPheLeuLeuIleHisGlyThrAlaAspAsnValHisPhe 661
 QY 2450 TTCACACAAACTTCCTCGTCTCCCACTGATCCGAGAGGAAACCTTACCAGCTCCAG 2509
 Db 662 GlnAsnSerMetGluPheSerGluAlaLeuIleGlnAsnLysGlnPheAspPheMet 681
 QY 2510 ATCTACCCCAAGAGAGACACAGTATT 2536
 Db 682 AlaTyrProAspLysAsnHisSerIle 690
 RESULT 5
 T41703
 dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T41703
 R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, August 1998
 A:Reference number: 222011
 A:Accession: T41703
 A:Status: preliminary; translated from GB/EMBL/DDBT
 A:Molecule type: DNA
 A:Residues: 1793 <MUR>
 A:Cross-references: EMBL:AL031180; PIDN:CAA20138.1; GSPDB:GN00066
 A:Experimental source: strain 972h-; cosmid c2E11 -chimeric
 C:Genetics:
 A:Gene: SPAC2E11.08
 A:Map position: 1
 C:Superfamily: dipeptidyl-peptidase IV
 Alignment Scores:
 Pred. No.: 1.16e-22 Length: 793
 Score: 508.50 Matches: 178
 Percent Similarity: 39.66% Conservative: 100
 Best Local Similarity: 25.39% Mismatches: 274
 Query Match: 10.60% Gaps: 149
 DB: 2 Gaps: 25
 US-09-976-674-4 (1-2617) x T41703 (1-793)
 QY 614 TCCTTCATCAATTAACAGCAGCTGTGGGTGCCAATCGAGAGCGAGAGCGGCG 673
 Db 184 SerPheValIyrAsnAsnAspLeuTyrValArgLys--AsnAspGlyAsnValGlnArg 202
 QY 674 CTGACCTTCGCCACCAAGGTTTATCCAAATCTCTGATGACCCCAAGTCTGGCGGTG 733
 Db 203 LeuThrTyr-----AspGlyThrValAspValPheAsn-----GlyLeu 215
 QY 734 GCCACCTTCGTCATACAGAAAGAG---TTCGACCGCTTCACACTGGGTACTGGTGGCC 790
 Db 216 ThrAspTrrPleIyrGluGluGlnValLeuSerSerProSerThrIleTrrPrrSerPro 235
 QY 791 ACAAGCTCTGGGAAAGTTCAAGAGGCGCTCAAGAGCTCGCAATCTGTATGAGAAAGTC 850
 Db 236 AspSerAsp-----LysIleAlaPheLeuLysLeu 245
 QY 851 GATGAGTCCGAGGTGAGGTATTCAGCTCC-----TTCCTCGG 892
 Db 246 AsnIleSerGlnIleProThrTyrHisIstYrProLeuTyrThrAlaGluLeuAspProSer 265
 QY 893 CTAGAAATA-----AGGAAGCGAGCTGATCGTATCGATCCCGAGAGCGAGAG 943
 Db 266 LeuProGluIurPheAspTyrAsnLysAspMetAlaIleLysTyrProLysProGluYAsnPro 285
 QY 944 AATCCCAAGATTCCTTGAATCGGTGAGTTCAGACTGACGACGAGCGCAAGATGTC 1003
 Db 286 AsnProSerValSerLeuPheValAlaAspLeuAsnSerAsnAlaSerSerAsnPheSer 305
 QY 1004 TCACACCAAGAGAGAGCTGTGTCAGCCCTTCAGCTCGCTTCGCCGAAGGTGAGATAC 1063
 Db 306 LeuTrrHisAsnGlnProValLeuAlaGluPro-----Val 316
 QY 1064 ATCGCCAGGCGCGGTGGACCCGCGGATGGCAATATGACCTGGGCCATGTTCTGAGCCG 1123

Db 317 ValGlnAsnValLeuTrrValAsnThrSerSer---ValLeuValGlnPheThrAsnArg 335
 QY 1124 CCCCAGAGTGGCTTCACAGCTGTCTCTCTCCCGGCGCTTCATCCAGACAGAG 1183
 Db 336 AsnSerThrCysIleThrAlaArgLeuLeuAspThrGluLysSerIleHisThrVal 355
 QY 1184 AATGAGGAG-----CAGGGCGTACGCTCGCCAGAGCGTGTCC 1222
 Db 356 LysThrGluCysLeuGluGluGlyTrrTyrGluValGlnGlnSerAlaLysMetPhePro 375
 QY 1223 AGGAATGTCCAGCCGATGTGCTGATGAGAGGCTCAACAGCTGTGATCAATGTTCAT 1282
 Db 376 LeuAsn-----AsnSerLeuValTrrGluAsnTrrSer 386
 QY 1283 GACATCTTCTATCCCTTCGCCAATCAAGAGGAGAGAGAGCTGTCTCTCGCGCC 1342
 Db 387 AspGlyTyrPheAspIleLeuAlaLeuAspAspTyrAsnHisLeuAlaPheIle----- 404
 QY 1343 AATGAATGCAAGACCGGCTTCGCAATTTGTACAAAGTCACCGCGTTTAAATCCAG 1402
 Db 404 ----- 404
 QY 1403 GGTACGATTTGAGTGAAGCCCTTCAGCCCGGAGAGATGAATTTAAGTCCCATTAAG 1462
 Db 405 -----ProPheAsnGly-----Ser 409
 QY 1463 GAACAGATGCTCTGACACGCGGTGAATGGAGTTTTGGCAGGACGCTCCAGATC 1522
 Db 410 SerProIleTyrLeuThrSerGlyAlaTrrAspValThr-----AspGlyProIle 426
 QY 1523 TGGTCAATGAGGAGACCAACAGCTGTGACTTCCAGGGCACCAAGAGACGCGCTGGAG 1582
 Db 427 HisIleAspGlyAspPheGlyAsnValTrrPheLeuAlaTrrLeuLysAspSerThrGlu 446
 QY 1583 CACCACCTTCAGTGTGATGAGCTATGAGGCGCGCGAGATCTGACGCTTACACAGCC 1642
 Db 447 ArgHisLeuTyrTrrValSerLeuAspThrLeu---GluIleTrrGlyIleThrAspAsn 465
 QY 1643 GCGTTCTCC-----CATAGCTGCTCCATGAGCCAGAACTTCGACATGTCGTACG 1693
 Db 466 GlnLysAspGlnGluGlyTrrTyrSerThrSerPheSerProPheGlyAspPheThrValLeu 485
 QY 1694 CACTACAGCAGCGGTGACAGCGCCGCTGCTGACGCTTACAGCTGAGCGGCGCCGAG 1753
 Db 486 AsnTyr-----HisGlyProAsp 491
 QY 1754 GACGACCCCTGCACAAGACAGCCCGCTGCGGTGATGATGAGAGCGACGCTGC 1813
 Db 492 ValPro-----TrrPheGluLeuAspSerThrLysAsp--- 502
 QY 1814 CCCCCGATTAAT----- 1825
 Db 503 ---LysAspTrrCysLeuSerLeuGluThrAsnSerArgLeuLysGlnLeuSerSer 521
 QY 1826 -----GTTCCTCCAGAGATCTTCATTTCCACAGCGGCTGCGGTGCGGCTTACGCG 1879
 Db 522 IleThrLeuProSerValGluTrrGlyLysLeuThrPheAsnAspThrThrPheAsnPhe 541
 QY 1880 ATGATCTTACAAAGCCCGCTTCAGAGCTGTCAGAGGAAAGACCCGACGCTCTTGTGA 1939
 Db 542 MetGluTrrArgTrrProAlaGlnPheAspValAlaAsnLysTyrTrrValLeuPhePheAla 561
 QY 1940 TATGAGGCGCCCGAGGTGACGCTGTGAATTAATCTTCAAAAGCATCAAGTACTTGG 1999
 Db 562 TrrGlyLysProGlySerGlnValAlaLysLeuPheArgValAspPheGlnAlaTyr 581
 QY 2000 CTCAACACACTGGCTCCCTGAGCGGTAGCGGCTGTGTGATGACGCGAGGCGTCTGT 2059
 Db 582 LeuAlaSerHisProAspPheGluPheIleValAlaThrLeuAspGlyArgGlyThrGly 601
 QY 2060 CAGCGAGGCGCTTGTGCAAGGGGCGCTGAAAAACCAATGGGCGAGGTGAGATCGAG 2119

Db 208 LAspArgHisThrGlyTyrTrpTrpAlaProAspSer----- 221
 Qy 820 CAAGACCCCTGCAATCTGTATGAGGAAGTATGATCCGAGGTGAGTCAATTCAGT 879
 Db 222 -----AlaIleuAlaTyrAlaArgIleAspGluSerProVal----- 235
 Qy 880 CCCCTCTCTGCGGTGAGAGAAAG-----AAGACGAC-----TC 915
 Db 236 -----GlnLysArgTyrGluValTyrAlaLysPargTrpAlaIleGlu 250
 Qy 916 GTATCGGTACCCCGAGACAGACAGAAATCCCAAGATTCCTTGAACCTGGCTGAGT 975
 Db 250 uGlnAArgTyrProAlaIleAspAlaAsnValGlnValLysLeuGlyAlaIleSerPr 270
 Qy 976 CCAAGATGACAGAGCCAGGAGCATGCTGACCCAGGAGAGAGAGGTGGTGCAGCCCTT 1035
 Db 270 oAlaGluGlnAlaGlnThrGlnTrpIleAspLeuGlyLysGluIleAspIle----- 287
 Qy 1036 CAGCTCGCTGTCCCGAAGGTGAGTACATCCGAGGCGCGGTGACCCGCGATGGCAA 1095
 Db 288 -----TyrLeuAlaArgValAsnTrp---ArgAspProGlu 298
 Qy 1096 ATAGCGCTGGGCGCATGTCTCTGGACCCGCGCCAGACAGTGGCTCCAGCTCGTCTCC 1155
 Db 298 nHisLeuSerPheGlnAArgGlnSerArgAspGlnLysLeuAspLeuVal----- 315
 Qy 1156 CCGGCGCTGTTCATCCCGAGACAGAGAATGAGGAGCAGGCGGTAGCTGTGCCAGAC 1215
 Db 316 -----GluValThrLeuAlaSerAsn----- 322
 Qy 1216 TGTCCCGAGAAATGTCAGCGGTATGTGTTGATGAGAGAGTCCAAAGCTGGATCAA 1275
 Db 323 -----GlnGlnAArgValLeuAlaHisGlnThrSerProThrTrpValPr 337
 Qy 1276 TGTTCATGACATCTTCTATGCTCCCTCCCAATCAGAGAGAGACAGACGCTCTCTCT 1335
 Db 337 oLeuHisAsn-----SerLeuAArgPheLeuAspArgLysSerIleLe 351
 Qy 1336 CCGCGCAATGAATGACAGACCGGCTTGTGATTTGATACAAATGTCACCGCGTTTAA 1395
 Db 351 uTrpSerSerGlu---ArgThrGlyPheGlnHisLeuTyrArgIle-----As 366
 Qy 1396 ATCCAGGCTAGCATTTGAGTGAAGTGAAGCCCTTACGCGCGGAGATGATTTAAGTCCC 1455
 Db 366 pSerLysGly----- 369
 Qy 1456 CATTAAGGAAGATTTCTGTGACACCGGTGAATG-----GAGTTTGGCGAG 1506
 Db 370 -----LysAlaAlaAlaLeuThrHisGlyAsnTrpSerValAspGluLeuAla-- 386
 Qy 1507 GCAGGCTCCAAAGATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGCGACCAA 1566
 Db 387 -----ValAspGluLysAlaGlyLeuAlaTyrPheAlaGlyIle 400
 Qy 1567 GGAACGCGCCCTGGAGACCAACCTTACGTGCTGAGCTATGAGCGCGCGAGATCGT 1626
 Db 400 eGluSerAlaArgLysSerGlnIleTyrAlaValProLeuGln---GlyGlyGlnProGlu 419
 Qy 1627 ACAGCTACACAGCCCGGCTTCTCCATAGCTGCTCATGAGCCAGAACTTGCACATGTT 1686
 Db 419 nArgLysSerLysAlaProGlyMetHisSerAlaSerPheAlaArgAsnAlaSerValTy 439
 Qy 1687 CGTACGACCTACAGCAGCGGTGAGAGACCAAGCTGGTGTACTTCCAGGCGACGAG 1746
 Db 439 rValAspSerTrpSerAsnSerTrpProProGlnIleLysLeuPheArgAlaAsnGlu 459
 Qy 1747 CCGCGAC-----GACGACCCCTTGCACAGACGACCGCG 1779
 Db 459 yGluLysIleAlaThrLeuValGluAsnAspLeuAlaAspProLysHisProTyrAlaAr 479
 Qy 1780 CTTCCTGGGCTAGCATGATGAGAGCAGCAGCTGCGCGCGGATTAATGCTCTCCAGAGAT 1839
 Db 479 gTyr-----ArgGluAlaGlnArg-----ProValGlu 488

Qy 1840 CTTCATTTTCCACAGCGGCTGGATGTGCG-----CTTACGCGCATGATCTACAGCC 1893
 Db 488 uPheGlyThrLeuThrAlaAlaAspGlyLysThrProLeuAsnTrpSerValIleLysPr 508
 Qy 1894 CCACGCTTGCAGCCAGGAGAAAGAACACCCAGCTCTCTTGTATATGAGGCCGCCA 1953
 Db 508 oAlaGlyPheAspProAlaLysArgTyrProValAlaValTyrGlyGlyProAl 528
 Qy 1954 GGTCCAGCTGGTAATACCTCTTCAAGGATC---AAGTACTTGGCGGTCAACACACT 2010
 Db 528 aSerGlnThrValThrAspSerTrpProGlyArgGlyAspHisLeuPheAsnGlnTyrLe 548
 Qy 2011 GGCCTCCCTGGCTACGCGCGGTGTGATGATGAGGAGGCGCTCTGTCAGGAGGCT 2070
 Db 548 uAlaGlnGlnGlyTyrValValPheSerLeuAspAsnArgGlyThrProAlaArgGlyAr 568
 Qy 2071 TCGGTTGCAAGGCGCTGAAACAAATGAGCCAGGTGAGATCGAGACAGGTGGA 2130
 Db 568 gAspPheGlyGlyAlaLeuTyrGlyLysGlnGlyThrValGluValAlaAspGlnLeuAr 588
 Qy 2131 GCGCTCGAGTCTGCGCGCGAGATGATGCTTATGATGAGCTGAGCCGATGCCATCCA 2190
 Db 588 gGlyValAlaTrpLeuLysGlnGln---ProTrpValAspProAlaArgIleGlyValGlu 607
 Qy 2191 TGCCTGCTTACGCGGCGCTCTCTGCTCATGAGGCTAATCCAAAGCCCGAGTGT 2250
 Db 607 nGlyTrpSerAsnLysGlyTyrMetThrLeuMetLeuLeuAlaLysAlaSerAspSerTy 627
 Qy 2251 CAACTGGCCATCCGCGGCGCGCGGTGACCGCTGATGGCTGATGAGACAGAGGTACAC 2310
 Db 627 rAlaCysGlyValAlaGlyAlaProValThrAspTrpGlyLeuTyrArgSerHisTyrTy 647
 Qy 2311 TGAAGCTTACATGAGCTCTCTGAGAACACACAGCGCTATGAGGCGGTCCGTGGC 2370
 Db 647 rGluAArgTyrMetAspLeuProAlaArgAsnAspAlaGlyTyrArgGluAlaArgValLe 667
 Qy 2371 CCTGCACTGGAGAGCTGCCCAATGAGCCCAAGCTGCTTATCCCTCCAGCGCTCT 2430
 Db 667 uThrHisIleGluLysLeuAlaArgSerPro-----LeuLeuLeuIleHisGlyMetAl 684
 Qy 2431 GGAAGAAACCTGCACTTTTTCACACAAATCTCTGCTCCCACTGATCCGAGCAG 2490
 Db 684 aAspAspAsnValLeuPheThrAsnSerThrSerLeuMetSerAlaLeuGlnLysArgGlu 704
 Qy 2491 GAAACCTTACAGCTCCAGATCTACCCCAAGAGACAGATGTTGCTGCTCCCGAGTC 2550
 Db 704 yGlnProPheGluLeuMetThrTyrProGlyAlaLysHisGlyLeuSerGlyAlaAspAl 724
 Qy 2551 GGGCGACCACTATGAAGTC 2569
 Db 724 aLeuHisArgTyrArgVal 730

RESULT 4
 S66261
 x-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
 C:Species: Flavobacterium meningosepticum
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
 C:Accession: S66261
 R:Kadashima, T.; Yoshida, T.; Ito, K.; Yoshimoto, T.
 Arch. Biochem. Biophys. 320, 123-128, 1995
 A:Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from
 A:Reference number: S66261; MIMD:95314307; PMID:7793970
 A:Accession: S66261
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-711 <KAB>
 A:Cross-references: EMBL:PD2121; NID:9577283; PIDN:BA07702.1; PID:9577284
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidyl-peptide hydrolase

Alignment Scores: 2.69e-27 length: 711
 Pred. No.:

```
Db 382 LeuGluGly-----ValAspGluAlaArgLysValAlaIlePheSerAlaSer 397
1565 AAGAGACGGCGGTGAGACACCACTGCTACGTCAGCATGAGGCGGCGGAGATC 1624
Db 398 IleAspThrProIleGluArgLeuTyrGluValSerTyrAlaLysProGlyLysPro 417
1625 GTAGCGCTCACACAGCGCGCTTCCCATAGCTGCTCCATGAGCAACTTCGACATG 1684
Db 418 LysAlaLeuThrSerIleGlyLysTyrPTrPTrAlaIleAlaLysAspGlyAla 437
1685 TTGTCAGACCACTACAGACGCTGAGACGCGCCCTGCTGACACTTCACAAAGCTGAGC 1744
Db 438 PheIleGlyThrTyrSerAspProLysThrProSerGlnThrAlaLeuTyrSerAlaAsp 457
1745 GGGCCC-----GACGAGACCCCTGTCACACAGACGCGCCCTTGAGGCT 1789
Db 458 GlyLysArgValArgTrpIleGluLysLysLysLeuAlaGluLysIleHisProTyrTrpPro 477
1790 AGCATGATGAGAGCAGCCAGCTGCGCCCGGATTAATGTTCCACAGAGATCTTCATTTC 1849
Db 478 -----TyrAlaIleAsnLeuPro-----GlnProGluPheGlySerLeu 490
1850 CACACGCGCTGGATGTGGGCTTACGGCATATCTACAGCCCAAGCCTTGACGCCA 1909
Db 491 LysAlaIleAspGlyLysThrLeuHisTyrGluIleLeuLysProIleGlyPheAspPro 510
1910 GGGAGAGACGCGCCAGCCGCTTGTATGAGAGCGCCGCGAGCGCGAGTGAAT 1969
Db 511 AlaLysLysTyrProAlaIleValSerValTyrGlyLysProHisAlaGlnArgValMet 530
1970 AACTCTTCAAGAGCATCAAGTACTTGGGCTCAACACAGCTGCGCTCCCTGGGCTAGCC 2029
Db 531 LysAsnTrpHisSerProSerGlu-----ArgThrTyrLeuGluAlaGlyTyrVal 547
2030 GTGCTGTGATTGACGAGCGGCTCCTGTCAGCAGAGGCTTGCTCGAAGGCGCCTG 2089
Db 548 IlePheLysLeuAspAsnArgGlySerGlyAsnArgSerAlaLysPheMetArgAlaLeu 567
2090 AAAAACCAATGGGACGAGATCGAGACGAGTGGAGTGGAGCGGCTGCGGCGC 2149
Db 568 AspArgLysLeuGlyThrValGluValGluAspGlnLeuGluLysAlaLysPheLeuAla 587
2150 GAGAGTATGCTTCATGCACTGAGCGGAGTGGCATCCATCCAGTGGTGGTCCAGGGGCG 2209
Db 588 SerGln---ProTyrValAspAlaAspLysLeuGlyValMetGlyTrpSerTyrGlyGly 606
2210 TTGCTCTCGCTTACGGGCTAATCCACAGCCCGAGCTGTTCAGAGTGGCCATCGGGGT 2269
Db 607 PheMetAlaLeuMetLeuLeuThrAlaGluAsnThrProPheLysAlaGlyAlaIleGly 626
2270 GCGCGGTCACGCTGATGAGGCTGACGACAGAGGATACAGTGGAGGCTGATGAGGAGCTC 2329
Db 627 AlaProThrThrLysTrpSerLeuTyrAspThrAlaTyrThrLysAlaGlyTyrMetGlyLys 646
2330 CCGTGAACAACACAGACGCGTATGAGGCGGGTCCGTGGCGCTGACAGTGGAGAACTG 2389
Db 647 ProAspGluAsnLysAlaGlyTyrAlaTyrSerAspLysAsnArgLysIleAspLysLeu 666
2390 CCCAATGAGCCCAACGCGTGTATTCCTGACAGGCTTCTGACGAGAAACCTTGACACTTT 2449
Db 667 -----AlaTrpGlySerLeuLeuLeuHisGlyMetAlaAspAsnValIlePhe 684
2450 TTCCACACAATCTCTCGTCCCAATGATCCGAGACGAGGAAACCTTGACAGCTGAC 2509
Db 685 GluAsnSerThrArgLeuMetAlaIleAlaLeuGlnArgLysAlaIleLeuPheGluMetAla 704
QY 2510 ATCTACCCCAACAGACAGACAGT 2533
Db 705 MetTyrProGlyLysThrHisSer 712
RESULT 3
JCS142
```

```
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C:Species: Xanthomonas maltophilia
C:Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C:Accession: JCS142
R:Kashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A:Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and express
A:Reference number: JCS142; MUID:97164011; PMID:9010758
A:Accession: JCS142
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-741 <KAB>
A:Cross-References: DDBJ:D83263; NID:g1753196; PIDN:BA11872.1; PID:g1753197
C:Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptidase hydrolase; membrane bound
F:4-18/Domain: transmembrane #status predicted
F:610/Active site: Ser #status predicted
F:685/Active site: Asp #status predicted
F:717/Active site: His #status predicted
```

Alignment Scores:

pred. No.:	1,03e-38	Length:	741
Score:	776.00	Matches:	238
Percent Similarity:	44.86%	Conservative:	133
Best Local Similarity:	28.78%	Mismatches:	295
Query Match:	16.18%	Indels:	161
DB:	2	Gaps:	32

US-09-976-674-4 (1-2617) x JCS142 (1-741)

```
QY 202 GTGTCGGCCCACTCCACGCGCTTACTGAGGAATGCAATATGACCGCAGAGA 261
Db 28 IleThrGlyProLeuPro-----LeuSerGlyProThrLeuMetLysProLysVal 44
QY 262 CTCCTCTCTACTGATGCAATTCC-----AACAGTCCGGAAGAGCTCT 309
Db 45 AlaProAsp-GlySerArgValThrPheLeuArgLysAspSerAspArgGlnLeu 64
QY 310 GCGTGCCTGCTCTGCG-----AACGACATGCTGATCATTTCCA 348
Db 64 uAspLeuTrpSerTyrAspIleGlySerGlyGlnThrArgLeuLeuValAspSerLysVa 84
349 GGCACGCGCCCAACATGAGGCTTACTGCGGAGAGAGAGCTCTGAGGAGCGGAACG 408
QY 84 ValLeuProGlyThrGluThrLeuSerAspGluGluLysAlaArgArgLysGlnThr 104
QY 409 CTTGGGGGCTTC---GGCATCACTCTGACACTTCACAGCGAGAGTGGCTTCTCT 465
Db 104 GlieAlaAlaMetThrGlyIleValAspTyrGlnTrpSerProAspAlaGlnArgLeu 124
QY 466 CTTCCAGGCGCAGACAGC-----CTCTTCACCTGCGGAGGCGGGAAGAGGCTT 519
Db 124 uPheProLeuGlyGlyLeuLeuTyrLeuTyrAspLeuLysGlnGlyLys----- 141
QY 520 CATGTGTCCCTATGAACCGCTGGAATCAAGACCCAGCTGCTCAGGCGCCGATGGA 579
Db 142 -----AlaAlaValArgLysLeu-----ThrHisGlyGlnGlyPheAlaThrAs 156
QY 580 CCCCAAAATGTGCGCTTCGCGACCTGCTTCTCTTCATGATAATTAACGCGACTGTG 639
Db 156 PAlaLysLeuSerProLysGly---GlyPheValSerPheIleArgGlyArgAsnLeuTr 175
QY 640 GGTGGCAATGAGAGCAGGAGGAGCGGCTGACCTTGTGCAACACAGATTATC 699
Db 175 PValIleAspLeuAlaSerGlyArgGlnMetGlnLeuThr-----AlaAspLysThr 193
QY 700 CAATGCTTGATGACCCCAAGTCTGGGGGTGCGCACCTTGCTGATACAGGAAGATT 759
Db 193 rThrIle-----GlyAsnGlyIleAlaGluPheValAlaAspLysGluMet 208
QY 760 CGACCGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 819
```

```

Db      781 HisGlyTyrPserTyrGlyGlyTyrMetAlaLeuGlnMetIleAlaLysHisProAsnIle 800
OY      2249 TTCAGAGTGGCCATCGGGGGGGGGCCCGGTGCACGGTCGATGGCTCGACGACAGAGGTAC 2308
Db      801 TTTATGAlaAlaIleAlaGlyGlyAlaValSerAspTrpArgLeuTyrAspThrAlaTyr 820
OY      2309 ACTGAGGCTACATGAGAGTCCCTGAGAACACAGCAGCGGCTATGAGGGGGTTCGGTG 2368
Db      821 ThrGlnArgTyrMetGlyTyrPro---LeuGlnGlnHisValTyrGlyAlaSerSerIle 839
OY      2369 GCCCTGCACCTGGAGAGAGTGGCCCAATGAGCCCAACGGCTTCTATCTCCACGGCTTC 2428
Db      840 ThrGlyLeuValGlnTyrLeuProAspGluProAsnArgLeuMetLeuValHisGlyLeu 859
OY      2429 CTGAGCAAAAACGTGACCTTTTCCACAAACACTCTCTCGTCCCGCAACGATCCGAGCA 2488
Db      860 MetAspGlnAsnValHisPheAlaHisLeuThrHisLeuValAspGluCysIleLysLys 879
OY      2489 GGGAAACCTTACACAGCTCCAGATCTACCCCAACGAGAGACACAGATATTCGCGCCGAG 2548
Db      880 GllYsrTrpHisGlnLeuValIlePheProAsnGlnArgHisGlyValAlaGAsnAsnAsp 899
OY      2549 TCGGGCGACGACTATGAGTCAAGCTGCGCTGCACTTTCTACAGAA 2593
Db      900 AlaSerIleTyrLeuAspAlaArgMetCysTrpPheAlaGlnGln 914

```

RESULT 2

AB87516
dipeptidyl peptidase IV [Imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

R:Accession: AB87516

R:Name: W.C.; Feldbylum, T. V.; Paulsen, I. T.; Nelson, K. E.; Eisen, J.; Heidelberg, J. B.; Laub, M. T.; Deboy, R. T.; Dodson, R. J.; Durkin, A. S.; Gwinn, M. L.; Haft, D. H.; Kolton, J.; Ermolova, M.; White, O.; Salzberg, S. L.; Shapiro, L.; Venter, J. C.; Fraser, C. M.; Proc. Natl. Acad. Sci. U.S.A. 98, 4116-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: AB87249; MUID:21173698; PMID:11259647

A:Accession: AB87516

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-738 <STO>

A:Cross-references: GB:AE005673; NID:g13423647; PIDN:AAK24125.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC2154

Alignment Scores:

```

Pred. No.:      1,74e-40      Length:      738
Score:          805.50        Matches:      227
Percent Similarity: 43.78%    Conservative: 118
Best Local Similarity: 28.81%  Mismatches:    308
Query Match:    16.80%       Indels:       135
Db:             2            Gaps:         25

```

US-09-976-674-4 (1-2617) x AB87516 (1-738)

```

OY      209 CCCACATCCACCGCCCTCTACTACCTG---GSAATGCCATATGGCAGCCGAGAACTCC 265
Db      47 ProAspGlyLysArgValThrTyrLeuLysGlyLysProGlnAlaAlaAsnIleGlnAsp 66
OY      266 CTCCTCTACTCTGAGATT-----CCCAAGAAAGTCCGGAAGAGGCTCTGCTG 313
Db      67 LeuTrpAlaAlaAspValLysGlyGlnProTyrArgLeuIleAspSerAlaAlaLeu 86
OY      314 CTCCTCTCTGAGAGCAGATGCTGATCATTTCCAGGCCACGCCCATGGGGCTTAC 373
Db      87 SerSerGlyLysArgGlnLeu----- 93
OY      374 TTCGCGGAGAGAGAGCTGCTGAGAGAGCGAAACGCTGGGGGCTTCGGCATCACCTCC 433
Db      94 SerGlnAlaGlnLysAlaArgValArgGlnAlaArgValSerAlaArgGlyIleValGln 113

```

```

OY      434 TACGATTTCCACAGCAGAGAGTGGCTCTCTCTTCCAGGCGCACAAACGCTTCCAC 493
Db      114 TyrSerTrpAspArgGlnGlyArgPheIleLeuValProLeuAspGlyAspLeuTyr--- 132
OY      494 TGTCCGACGGCGGCAAGAACGCTTCATGGTGTGCCCTATGAACCGCTGGAATCAAG 553
Db      133 ---LeuAspAlaValAlaAspGly-----LysIleThrArgLeu 144
OY      554 ACCGAGTGTCAAGGGCCCCGGATGACCCCAAAATCTGCCCTGCCACCCCTTCTTC 613
Db      145 ThrGlnThrProGlyAspGlnValAlaSpAlaLysValSerProLysGly---GlyTyrVal 163
OY      614 TCTTCATCATATACAGCAGCAGCTGTGGTGGCCCAATCGACAGACAGCAGAGCGCGG 673
Db      164 SerTyrValArgAspGlnAsnLeuTyrIleLysProValAlaGlyGlyAlaGlnThrAla 183
OY      674 CTGACCTTCTGCCACCAAGCTTATGCCAATGCTCGATGACCCCAAGTCTGCGGTGTG 733
Db      184 LeuThrThrAspGlyLys-----AspAlaLeuSerPheGlyVal 196
OY      734 GCCACCTTCGTCATACAGAAAGAGTTCGACCGCTTCACTGGGTACTGTGTGCCCCACA 793
Db      197 AlaGlnPheIleValGlnGlnGlnLeuAspArgPheThrGlyTyrTrpSerProAsp 216
OY      794 GCTCTCTGGGAAAGTTCAGAGGGCTCAAGACGCTGCGAATCTGTATGAGAAAGTCAT 853
Db      217 GluSer-----ArgIleValTyrThrArgValAsp 226
OY      854 GAGTCCGAGGTGAGGTTCATTCACGTCCCTCTCTCGCTAGAGAAAGAGACGAC 913
Db      227 GluSerGlyValAspIleValProArgAlaAspIleGlyProGlyGlyAlaThrValVal 246
OY      914 TTCGATGGGATACCCGACGACAGCAGCAGCAAGAACCCCAAGTGGCTGAAACCTGCGAG 973
Db      247 AsnGlnTrpArgProAlaGlnArgProAsnAlaValAlaLysLeuPheValAlaArgAsp 266
OY      974 TTCAGACTGACAGCCAGCGCAGAGATGCTCGACCCAG-----GAGAAGAGACTG 1024
Db      267 Leu-----AlaSerGlyLysValThrAlaLeuAspLeuGlyAlaAlaAsnLysAspIle 283
OY      1025 GTGCAGCCCTTCACGCTCGCTTCCCGAAGGTGAGTACATCCGCCAGCGCCGGGTGACC 1084
Db      284 -----TyrValAlaArgValAlaIleTrpSer 291
OY      1085 CGGATGGCAATATACGCTGGGCCATGCTTCGTGACCGCGCCAGCAGTGGCTCCACTCC 1144
Db      292 AlaAspGlyLysThrValTyrValGlnArgLeuSerArgAspGlnLysThrLeuAspLeu 311
OY      1145 GTCTCTCTCCCGCGGCTGTTCATCCGACGACAGAGATGAGAGCAGCGCTAGCC 1204
Db      312 LeuAlaPheAspAlaIle-----ThrGly 319
OY      1205 TCTGCAGAGCTGTCCCGAAGATGTCACCGCTATGTGTGACGAGAGGTCCACCAC 1264
Db      320 AlaGlyLysThrIleLeuThrAspThrAspProHisPheIle-----GluValSerAsn 337
OY      1265 GTCTGATCATATGTATCATGATCTTATCCCTTCCCAATCAAGAGAGAGAGACGAG 1324
Db      338 AspHe-----ArgProLeuThrAspGly----- 345
OY      1325 CTCGTCTTCTCCGCGCAATGATGACAGACGCGCTTGCACATTTGTCAAAAGTCACC 1384
Db      346 ---ThrPheLeuTrpGlySerGln---LysAspGlyAsnGlnHisLeuTyrArgTyrAla 363
OY      1385 GCCGTTTAAATCCAGGCTACGATTGAGTGAAGCCCTTCAGCCCGGGGAAGATGAA 1444
Db      364 Ala-----AspGly----- 366
OY      1445 TTTAAGTCCCATTAAGAGAGATTTGCTTACACAGCGGTGAATGGAGGTTTGGCG 1504
Db      367 -----LysLeuIleAlaGlnIleThrLysGlyAspTrpProValIleGly 381
OY      1505 AGCAGCGCTCCAAAGATCTGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGGCACC 1564

```


QY 320 TCCTGGAAACAGATGCTGATCATTTCCAGGCCAGCCCCACCATTGGGGTCTACTCTCGG 379
||| : : : : :
Db 122 SerGIYTyAsnValAspSerTyrIleArgMetSerCysArgLysThrProProSerAla 141
QY 380 GAGGAGGAGCTCTCTGAGGAGCGGAAACGCTGGGGGTCTTC--GGATCACTCTCTAC 436
||| : : : : :
Db 142 GlupheThrLeuGlnCysGlnArgGlnArgSerGlnValValThrcdIYIleSerAspTyr 161
QY 437 GACTTCCACAGGAGAGTGGCTCTCTCTCCATCCAGGACGACGACGCTCTTCCAC-- 493
: : : : :
Db 162 GluIle-----ArgasnGlnLysMetIleLeuMetAlaGlnAspGlnLeuPheArgTyr 179
QY 493 ----- 493
Db 180 AsnProLeuAsnGlnAlaLeuAlaAlaIleProIleAlaValProAspAspGlnSerSer 199
QY 494 -----TGTCCGAC 502
Db 200 ThrGluProMetAspIleSerGlnGlySerIleThrSerGlyThrLysGlnCysSerAsn 219
QY 503 GCGCGCAACAGCGCTTCATGGTGCCTATGAACCCCTGGAATCAAG----- 553
Db 220 GluAlaProGlnSerSerThrValProProValThrArgIleProIleLysLysProThr 239
QY 554 ACCCACTGCTCAGGGCCCGATGACCC-----AAA 586
||| : : : : :
Db 240 ThrSerThrGlnLysProAlaThrAlaProProThrAsnAsnPheValSerSerAlaLys 259
QY 587 AATCGCCCTCGGACCCCTCTCTCTCTCTATCAATACAGCGACCTGGGGGTGCC 646
: : : : :
Db 260 ValCysProAlaAspSerSerLeuLeuAlaTyrValLeuAsnLysGlnValTyrIle-- 278
QY 647 AACATCGACAGAGCGGCGGCTGACCTTCTGCCACCAAGGTTATCAATATGC 706
279 -----GluLysasnGlnLysIleIleHisArgThrSerSerAsn-- 291
QY 707 CTGGATGACCCCAAGTCTCGGGTGTGGCCACCTTCATACAGAGAAGAGTTCCAGCCG 766
||| : : : : :
Db 292 ---SerLysHisIleThrAsnGlnValProSerTyrIleValGlnGlnGlnLysGlnLys 310
QY 767 TTCCTAGGCTACTGGTGTGGCCACAGCTCTCGGGAAGTTCCAGAGGCGCTCAAGACG 826
||| : : : : :
Db 311 PheGlnGlyIleThrPrr-----SerGlnSer--LysThr 321
QY 827 CTGCGAATCTGTAGAGAGTGCATGCATGCCAGTGGAGGTC-----ATT 874
||| : : : : :
Db 322 ---ArgLeuLeuTyrGlnHisValAsnGlnGlnLysValAlaGlnSerGlnPheGlnVal 340
QY 875 CACGTCCCTCTCTCTGCTAGAAAGAAAGAGACGACTGTATGCTGATCCCAAGACA 934
: : : : :
Db 341 AsnGlnAspProProValAla-----ProMetLysTyrProAlaGln 354
QY 935 GCGACGACAAGATCCCAAGATTGCTGAAATGGCTGATCCAGATCCAGACGAGGC 994
||| : : : : :
Db 355 GlyThrLysAsnAlaTyrSerThrLeuArgMetValIleLeuGln-----AsnGly 371
QY 995 AAGATGCTGTCCAGCCAGAGAGAGCTGGTGGACACCTTCAGTCGCTGTTCGCGAAG 1054
||| : : : : :
Db 372 LysAlaTyrAspValProLeuLysAspGlnVal-----IleTyrLysHisCysProPhe 389
QY 1055 GTGGAGTATACATCCAGCGCGGTGGACCGGAGTGAATAATACGCTGGCGCATGTTC 1114
||| : : : : :
Db 390 TyrGlnTyrIleThrArgAlaGlnPhePheSerAspGlyThrThrValThrValGlnVal 409
QY 1115 CTGGACCGGCGCCAGACGATGGCTCAGCTGCTCTCTCTCCCGCGGCTGTTCATCCG 1174
: : : : :
Db 410 MetSerArgAspGlnAlaGlnCysSerLeuLeuIleProTyrThrAspPheLeu 429
QY 1175 AGCAGCAGG----- 1183
Db 430 ProGlnGlnLeuGlnGlySerIleLysGlnAspAsnLeuGlnLeuSerThrAspLeuAsn 449

QY 1184 -----ATGAGAGACGAGCGGCTAGCGCTTGCACAGCTGTCCAGG----- 1225
Db 450 MetGlnValThrAspAspSerHisGlnGlnIleThrMetGlnLysProProArgGlnLys 469
QY 1226 -----AATGCCACCGCATGTGGTACGAGGAGGATCCACACGCTGGATCAT 1276
||| : : : : :
Db 470 LeuArgGlyThrValGln-----IleHisLysAlaArgAsnAspTyrThrIleAsn 486
QY 1277 GTTCATGACATCTCTATCCCTTCCCAATCGAGGGAGGAGACGACTCTGC--TTT 1333
||| : : : : :
Db 487 ThrHisAsnAlaIleTyrThrProLeuLysIleThrAspGlnGlnHisProMetCysGlnPhe 506
QY 1334 CTCCGCGCCCAATGAATGCAGACCGGCTTCTCCATTTGTACAAATGCACGCGCTTTTA 1393
: : : : :
Db 507 IleTyrCysLeuGlnLysProAsnGlnSerCys--LeuAlaLeuIleSerAlaGlnLeu 525
QY 1394 AAATCCAGGCTACGATTTGAGTAGACCCCTTACGCCCGGGGAAGATGAATTAAATGCC 1453
||| : : : : :
Db 526 AspGlnAsnGlyTyr-----Cys 531
QY 1454 CCCATTAGGAGAGATTTGCTGTACACGCGGTGAATGGAGGTTTGCGAGCGACGCG 1513
||| : : : : :
Db 532 ArgHisThrGlnGlnLysLeuLeuMetAlaGlnAsnPheSerIleAsnLysSerMetGly 551
QY 1514 TCCAAATCTGGGTCAATGAGAGACCAAGCTGTATCTTCCAGGACCAAGAGACAG 1573
||| : : : : :
Db 552 -----IleValValAspGlnValArgGlnLeuValTyrTyrValAlaAsnLysHis 569
QY 1574 CCGCTGGAGACCCACCTCTACGCTGACGTATGAGGCGCGCGAGATGCTACGCTTC 1633
||| : : : : :
Db 570 ProThrGlnTyrAsnIle---CysValSerHisTyrArgThrGlnHisAlaGlnLeu 588
QY 1634 ACCAGCCCGGCTCTCCATAGCTGCTCATGACGACCAAGC----- 1675
||| : : : : :
Db 583 ThrGlnSerGlyIle-----CysPheLysSerGlnAlaGlnLysLeuAla 605
QY 1676 TTCGACATG-----TTCGTACGCCATACAGACGCTGAGACGCGGCTGCTGT 1726
||| : : : : :
Db 606 LeuAspLeuAsnHisGlnLysPheAlaCysTyrMetThrSerValGlnSerProAlaGlnLys 625
QY 1727 CACGCTTACAACTGAGCGGCGCCGACGACGACCCCTGTGACAAACAGCCCGCTTCTG 1786
||| : : : : :
Db 626 ArgPheTyrSerPheArgThrLysGlnLysGlnValLeu-----ProSerThrValTyr 643
QY 1787 GCTAGCATGATGAGGACGACGACGCTGCCCC-----CCGAT-----TATGCTCTCA 1834
||| : : : : :
Db 644 AlaAlaAsnIleThrValSerGlnHisProGlnGlnProAspLeuHisPheSerPro 663
QY 1835 GAGATCTTCCATTTCCACAG--CGCTCGATGTGCGGCTTACGCGCATGATTAAG 1891
||| : : : : :
Db 664 GluMetIleGlnPheGlnSerLysThrGlnLysPheMetHisTyrAlaMetIleLeuArg 683
QY 1892 CCCCACGCTTTCAGCCAGGAGAAAGACACCCACGCTCTTTGTATATGAGGCCCC 1951
||| : : : : :
Db 684 ProSerAsnPheAspProTyrLysTyrProValPheHisTyrValTyrGlnLysPro 703
QY 1952 CAGGTGACGCTGTGAATATCTCTCAAGGACATCAAGATCTGGGTCAACACAGT 2011
: : : : :
Db 704 GlyIleGlnIleValHisAsnAspPheSerThrIleGlnTyrIleArg-----Phe 720
QY 2012 GCGCTCTGGGCTAGCGCGCTGTGTGATTCAGCGAGGCGCTCTGACGAGGAGGCT 2071
||| : : : : :
Db 721 CysArgGlnGlyTyrValValAlaValPheIleAspAsnArgGlnSerAlaHisArgGlyIle 740
QY 2072 CCGTTGAAAGGCGCCCTGAAACCAATGGGCGCAGGTGGAATGAGAGCAGGTGAG 2131
||| : : : : :
Db 741 GlupheGlnArgHisIleHisLysMetGlyThrValGlnValGlnAspGlnValGln 760
QY 2132 GCGCTGACGCTGTGGCGGAGATAT---GGCTTCATGACCTGCGGAGTTCGATC 2188
||| : : : : :
Db 761 GlyLeuGlnMetLeuAlaGlnArgThrGlnGlyPheMetAspMetSerArgValValVal 780
QY 2189 CATGGCTGTCTACGCGGCGCTCTCTGCTCATGCGGCTTAATCCACAGCCCAAGGTG 2248

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 12, 2002, 11:43:57 ; Search time 40.5 seconds
(without alignments)
12423.890 Million cell updates/sec

Title: US-09-976-674-4
Perfect score: 4795
Sequence: 1 caagcttaccatgcgcaccca.....cttgagcgccgcgagtcgcg 2617

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_n2p.model -DBV=tblh
-Q=/gnu2.1/USPTO.spool/US0997674/runtc_04122002_162400_6009/app_query.fasta_1.2759
-DB=PIR_73 -OFMT=fastan -SUFFIX=ipr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.ccd1 -LIST=45
-DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=prco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0997674_gcen_1.1.44 @runtc_04122002_162400_6009 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	934.5	19.5	931	2	T32919
2	805.5	16.8	738	2	A87516
3	776	16.2	741	2	JCS142
4	585.5	12.2	711	2	S66261
5	508.5	10.6	793	2	T41703
6	487.5	10.2	766	1	CDH026
7	484.5	10.1	792	1	A39914
8	480	10.0	760	1	S23752
9	479.5	10.0	799	2	T25174
10	467.5	9.7	779	2	T25173
11	452	9.4	865	2	T54331
12	445.5	9.3	803	2	I68600
13	442.5	9.2	803	2	A41793
14	432.5	9.0	818	1	A30107

15	417	8.7	931	2	A49737	dipeptidyl aminope
16	405.5	8.5	795	2	F82858	dipeptidyl-peptida
17	359.5	7.5	1367	1	F48478	glucan 1,4-alpha-g
18	352	7.3	2232	2	T34434	hypothetical prote
19	333.5	7.0	829	2	T19514	hypothetical prote
20	328	6.8	3570	2	T45025	mucin MUC5B, trach
21	314	6.5	759	2	T03853	fibroblast activat
22	302	6.3	1791	2	T02345	hypothetical prote
23	299	6.2	580	2	T43481	probable mucin DKF
24	291	6.1	4776	2	E95206	cell wall surface
25	286.5	6.0	683	2	E87495	prolyl oligopeptid
26	286	6.0	528	2	T47141	gastric mucin (clo
27	281.5	5.9	3020	2	A43932	mucin 2 precursor,
28	280	5.8	1106	2	T00405	hypothetical 119.5
29	276.5	5.8	924	2	S27923	gene lrf3 protein -
30	275.5	5.7	709	2	B82880	alanyl dipeptidyl
31	269.5	5.6	2187	2	T30826	nascent polypeptid
32	267	5.5	1414	1	S23809	collagen alpha 2(I
33	265.5	5.5	660	1	Q0BE3	BHLFI protein - hu
34	263	5.5	660	1	Q0BE3	BHLFI protein - hu
35	262	5.5	743	2	T37700	probable dipeptidyl
36	256.5	5.3	1032	2	T34433	hypothetical prote
37	255.5	5.3	1460	1	EDBEIF	immediate-early pr
38	254	5.3	1459	2	T32271	hypothetical prote
39	252	5.3	1777	2	T34369	hypothetical prote
40	252	5.3	13288	2	T03099	mucin, submaxillar
41	251.5	5.2	1763	2	S16366	collagen alpha 2(I
42	249.5	5.2	1690	1	CGH01B	collagen alpha 4(I
43	249	5.2	657	2	E70025	probable acylamino
44	248.5	5.2	839	2	F75518	hypothetical prote
45	248	5.2	1151	2	T18535	high molecular mas

ALIGNMENTS

RESULT 1
T32919
hypothetical protein K02F2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32919
R:Magdi, L.; Goela, D.
A:Submitted to the EMBL Data Library, January 1998
A:Description: The sequence of C. elegans cosmid K02F2.
A:Reference number: Z21246
A:Accession: T32919
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-931 <MAG>
A:Cross-references: EMBL:AF043699; PIDN:AA897564.1; GSPDB:GN00019; CESP:K02F2.1
A:Experimental source: strain Bristol N2; clone K02F2
C:Genetics:
A:Gene: CESP:K02F2.1
A:Map position: 1
A:Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2;

Alignment Scores:

Pred. No.: 3.47e-48
Score: 934.50
Percent Similarity: 45.36%
Best Local Similarity: 29.50%
Query Match: 19.49%
DB: 2
Length: 931
Matches: 264
Conservative: 142
Mismatch: 322
Indels: 167
Gaps: 33

US-09-976-674-4 (1-2617) x T32919 (1-931)
QY 224 CTCTACCTACCTGGGAATGCCATATGGCAGCCGAGAGACTCCCTCTACTCTGAGATT 283
DB 82 Metcylalalieserservalproglythrasmthcinsertilepheservalthrlle 101
QY 284 CCCAGAGAGCTC-----CGGAAGAGAGCTCTGCTCTCTG 319
DB 102 Proleuclnuuvalgluylsalaglinalalasparglysrhhegluueulsluyls 121

Run on: December, 12, 2002, 10:36:52 ; Search time 55 Seconds

(without alignments)
12680.614 Million cell updates/sec

Title: US-09-976-674-4

Sequence: 1 caagcttaccatgcca...tctgagcgccgcgatccg 2617

Scoring table: BLOSUM62

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

```

MODEL=irameat_nzp.model -DEV=xlh
-O/-cgrn2.1/USPO.spool/US099976674/runtat.04122002.162358.5964/app.query.fasta_1.2759
-DBA=Genesecd.101002 -QPM=fastan -SUFFIX=tag -MISMATCH=0.1 -LODCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsnum62 -TRANS=human0.0.cd1
-LIST=45 -DOCALLIG=200 -THR SCORE=pct -THR MAX=10 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -OUT=ext -HEAPSIZ=5E-50 -MINLEN=0 -MAXLEN=2000000000
-USER=US09976674.cgcN.1.1.61 -Eunit HBAS12002.162358.5964 -NCPUP=6 -ICP=3
-NO.XLXPY -NO.MMAP -LARGENQUERY -NEG_SCORES=0 -WAT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAEXT=7
YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : A_Geneseq_101002:*

1:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1980.DAT *
2:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1981.DAT *
3:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1982.DAT *
4:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1983.DAT *
5:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1984.DAT *
6:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1985.DAT *
7:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1986.DAT *
8:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1987.DAT *
9:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1988.DAT *
10:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1989.DAT *
11:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1990.DAT *
12:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1991.DAT *
13:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1992.DAT *
14:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1993.DAT *
15:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1994.DAT *
16:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1995.DAT *
17:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1996.DAT *
18:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1997.DAT *
19:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1998.DAT *
20:	/SID2/gcgdata/genseq/genseqp-emb1/AAL1999.DAT *
21:	/SID2/gcgdata/genseq/genseqp-emb1/AAL2000.DAT *
22:	/SID2/gcgdata/genseq/genseqp-emb1/AAL2001.DAT *
23:	/SID2/gcgdata/genseq/genseqp-emb1/AAL2002.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Query Score	Match Length	DB	ID	Description
---------------	----------------	-----------------	----	----	-------------

1	4646	96.9	863	23	ABG61592	Human DPPIV relate
2	4646	96.9	892	23	ABG61602	Human DPP-2 splic
3	4646	96.9	892	23	ABG61604	Human DPP-2 splic
4	4636	96.7	969	23	AAE24168	Human dipeptidyl p
5	4558.5	95.1	879	23	ABG61607	Human DPP-2 splic
6	4558.5	95.1	879	23	ABG61608	Human DPP-2 splic
7	4478	93.0	830	23	AAE24171	Human dipeptidyl p
8	4279	89.2	869	23	AAE24169	Alternative versio
9	4259.5	88.6	832	23	ABG61605	Human DPP-2 splic
10	4259.5	88.8	832	23	ABG61606	Human DPP-2 splic
11	4172	87.0	819	23	ABG61609	Human DPP-2 splic
12	4172	87.0	819	23	ABG61610	Human DPP-2 splic
13	4129	86.1	847	23	AAE23875	Human DPP-2 splic
14	3817.5	79.6	737	22	AAE23875	Murine dipeptidyl
15	3539.5	73.8	683	22	AAAM3724	Human polypeptide
16	3409.5	71.1	720	21	AAAB41626	Human polypeptide
17	2870	59.9	882	22	AAAB41780	Human OREF ORF1390
18	2870	59.9	882	22	AAE241717	Human DPP8, Homo
19	2870	59.9	882	23	ABG61591	Human dipeptidyl p
20	2870	59.9	882	23	AAU747249	Human DPPIV relate
21	2870	59.9	882	23	AAAG78415	Human protease PRRT
22	2688	56.1	497	23	ABG61645	Amnio acid sequenc
23	2688	56.1	497	23	AAU961619	Human albumin fusi
24	2684	56.0	497	23	AAU961619	Human secreted prote
25	2684	56.0	497	23	AAU961619	Human albumin fusi
26	2572	53.6	508	21	AAAB42928	Human secreted prote
27	2562	53.4	518	21	AAAG78415	Human OREF ORF2692
28	2547	53.1	518	21	AAV90299	Human DPP-2 splic
29	2414	50.3	782	23	ABG97361	Human peptidase, H
30	2236.5	46.6	512	22	AAAB40526	Novel human protei
31	2215.5	46.2	460	22	AAAM38740	Human polypeptide
32	2178	45.4	724	23	ABB97362	Human polypeptide
33	2137.5	44.6	632	22	AAAB95565	Novel human protei
34	2001	41.7	690	23	ABG61594	Human protein sequ
35	1999.5	41.7	661	23	ABG61596	Human DPP-1 splic
36	1997	41.6	658	23	ABG61600	Human DPP-1 splic
37	1816	37.9	613	23	ABG61601	Human DPP-1 splic
38	1631.5	34.0	1042	22	ABB60137	Drosophila melanog
39	1631.5	34.0	1102	22	ABBB6029	Drosophila melanog
40	1599.5	33.4	580	22	AAEL14337	Human protease PRRT
41	1494.5	31.5	465	22	AAAB47189	Human DPP8 318Thr
42	1220.5	25.2	360	22	AAAB47190	Human DPP8 244Glu
43	1093.5	22.8	310	22	AAAB47188	Human DPP8 524Phe
44	1093.5	22.8	310	23	ABBO8994	Human dipeptidyl p
45	1067.5	22.3	349	23	ABG64842	Human albumin fusi

ALIGNMENTS

RESULT 1

ID ABG61592 standard; Protein; 863 AA.

AC ABG61592;

DT 12-AUG-2002 (first entry)

DE Human DPPIV related serine protease DPPR-2.

KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP; DPP-IV

KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;

KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;

KW metabolic disorder.

05 Homo sapiens.

XX	
PN	WO200231134-A2.
PD	
XX	18-APR-2002.
XX	
PE	12-OCT-2001; 2001WO-US31874.
XX	
PR	12-OCT-2000; 2000US-240117P.
XX	
PA	(FERR) FERRING BV.
XX	
PI	Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX	
DR	WPI: 2002-444178/47.
XX	
DR	N-PsDB; ABR63323.
XX	
XX	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT	the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT	viral infections, cancers, allergies, neurological disorders, or pain
XX	
PS	Claim 17; Fig 1; 113pp; English.

The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPPIV)-related proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), pain, diabetes, precocious puberty, infertility, obesity, anorexia, bulimia, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, stroke, ulcers, asthma, allergies, cancers, migraines, vomiting, psychotic and neurological disorders (e.g. anxiety, dementia, or schizophrenia), and dyskinesias. These may also be used in discovering therapeutic agents for the treatment of reproductive, inflammatory or metabolic disorders. Ab651591-Ab651612 represent human DPPP proteins.

SQ Sequence 863 AA;

Alignment Scores:	
Pred. No.:	0
Score:	4646.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	96.89%
DB:	23
Length:	8653
Matches:	8633
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-976-674-4 (1-2617) x ABG61592 (1-863)

QY	11	ATGGCCACACACGGGACCCCAACGGCCGAGCAGCAGCAGCCGACAGATGACCCG	70
Db	1	MetalaethrhrhrglthrrprothhAlaspragblYuspralaalaatprAsprto	20
QY	71	GCGCGCCGCTTACAGGTGCAGAAACACTCGTGGACGGGCTCCGAGACATATCCAGGC	130
Db	21	AlaAlaArpRheglInValGlnLysHisSerTtrpAspLysLeuArgSerIleIleHisGly	40
QY	131	AGCGCCAAAGTCTGGGCGCTCATGTGTATACAAAGGCGGCCACGACTCCAGTTTGATG	190
Db	41	SertrglYusrYrSerLysLeuIleValAsnLysAlaArpHisAspRheglInpHeValGln	60
QY	191	AACAGCATGATGCTGGGCCCCACTCCACGCGCTCTACTACCTGGAAATGCATATGCG	250
Db	61	LysThrAspGluSerLysTyrProHisSerHisArgLeuYrLysLeuGlyMetProYrGly	80
QY	251	AGCGGAGAGAACTCCGCTCTACTCTGGAATTTCCCAAGAAGGCCGGAAGAAGGCTCG	310
Db	81	SertrglLusnsSerLeuLeuYrSerGluIleArpLysValArgLysGlnAlaLeu	100
QY	311	CTGCTCTGTCTCGGAAGCAGATGCTGGATCATTTCCAGGCCAGGCCCCACCATGGGGTC	370
Db	101	LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnIleArTrpHisLysGlyVal	120

[illegible]

671 CGGCTGACCTTCTGCCACCAAGGTTATCCAATGTCTGGATGACCCCAAGTCTGGGGT 730

Db 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240

731 GTGGCCACCTTCGTCATACAGGAGTTCGACCGCTTCACCTGGGTACTGGTGCCCC 790

Db 241 VALAIAthRphevalIleGInGIuGluPheaspargpetherGlyTYRTrpTrpCysPro 260

791 ACAGCCTCCTGGGAGGTTCAAGAGGGCCTCAAGACGCTGCGAATCCTGTATGAGGAAGTC 850

Db 261 ThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluVal 280

851 GATGAGTCCGAGGTGGAGGTCAITCAGGTCCCTCTCCTGCGCTAGAGAAGGAAGACG 910

Db 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArgLysThr 300

911 GACTTCGTAICGGTACCCGAGGACAGGACGCAAGATCCCAAGATTCCTTGAACCTGGCT 970

db 301 aspserlyrargyrproarg:inglyserllysasnprolysilealaleullysleuala 320

9/1 GAGTTCAGACTGACAGCCAGGGCAAGATCGTCTCGACCCAGGAGAAAGAGCTGGTGCAG 103

DD 321 GluPheGlnTyrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340

1031 CCCTTCAGCTTCGCTCTGTTCCCGAAGGTCGAGTACATCGCCAGGGCCGGTGGACCCGGAT 109

Db 341 ProhneserSerLeuphepPolysvalGluIYrIlealaargAlaGlyIrpIhrrargasp 360

QY 1091 GGCAATACGGCTGGGCAATGTCCTGGACCGGCCGACGATGGCTCCAGCTCGCTCTC 113

DD 301 GYLYS IYATAI PHAWELFNELEUASPAIGFIOWINGNITPLEUGINLEUVALLEU 300

1131 CCCCCCGGCTGTCAICCCGAGCACAGAAAGAGGAGCAGCGCTAGCCTCTGCC 121

DD 381 LEUPROPIOLALDEURNEI LEPROSERTINGUASHNGUGUGINHARGLEUADASEFALA 400

27 1211 AGAGCTGTCCTCAGGAATGTCAGCCGATGTTGGTGTACGAGGAGGTCACCAACGCTCG 127

401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

QY 12/1 AICAMGICATGACATCTCTAICCCCTCCCCCAATCAGAGGAGAGGACGAGCTCTGC 133

421 lrealsivalmsasprirefneiyiprofiereoginserguglyglaaspralutecys 440

QY 1351 TTTCTCCGCCAAATGAAATGCAGACCGGTCTCTGCCATTGTACCAAGTACC6CCGT 139

DD 441 FILED IN GALDASINGLUCUS LYSTINGLY PNECUSHLSLEU YL LYSVALTILDAVAL 460

QY 1391 TAAAHAIICCAAGGCIACGAIIGAGTIGAGCCCTTCAGCCCCGGGGAAGATGAATTTAAG 143

401 Leulysseerginglytyiaspiipsergluipriophneserproglygluaspluipheyls 480

QY 1451 TGCCCATTAAGAGAGATTGCTCTGACACAGCGGTGAATGAGAGATTGTCGAGAGCAC 1510
 |||||||
 Db 481 CysProLeuYsgLuGlulLeuHleuHrSerGlyLurPGLuValLeuAlaArgHis 500
 QY 1511 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGGTACTTCACAGGACCAAGAGAC 1570
 |||||||
 Db 501 GlySerYsIeTrrPValAsnGluGluThrYsLeuValTYrrPheGlnGlyThrLysAsp 520
 QY 1571 ACGCGCGGAGACACACCTTACGTGGTCACTATGAGGGGCGGCGGAGATGTTACGC 1630
 |||||||
 Db 521 ThrProLeuGluHISHisLeuTYrrValValSerTYrrGluAlaAlaGlyGluLeuValArg 540
 QY 1631 CTCACACAGCGCGCTTCTCCATAGCTGTCTCAAGACCAAGCTTCGACATGCTTCGTC 1690
 |||||||
 Db 541 LeuThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
 QY 1691 AGCCACTACAGACAGCTGAGACAGCGCGCTGCGTGCACGTCTACAGCTCAAGCGGCGCC 1750
 |||||||
 Db 561 SerHisTYrrSerSerValSerThrProProCysValHisValTYrrLysLeuSerGlyPro 580
 QY 1751 GAGACAGACCGCGTCGACACAGCGCGCTTGGGCTACCATGATGAGAGACCGCACG 1810
 |||||||
 Db 581 AspAspAspProLeuHISHisLysGlnProArgPheTrpAlaSerMetMetGluAlaAlaSer 600
 QY 1811 TGCCCGCGGATATATGTTCTCCAGAGATCTTCATTTCCACAGCGCGCTCGATGTGCGG 1870
 |||||||
 Db 601 CysProProAspTYrrValProProGluLePheHISrPheHISrHrArgSerAspValArg 620
 QY 1871 CTCTACGCGCATGATCTACACAGCGCGCTTCGACAGCGGAGAAAGACCGCACCGCTC 1930
 |||||||
 Db 621 LeuTYrrGlyMetIleTYrrLysProHisAlaLeuGlnProGlyLysHisProThrVal 640
 QY 1931 CTCTTTGATATGAGAGCGCGCGCTGAGTGTGAATCTCTTAAAGCATCAAG 1990
 |||||||
 Db 641 LeuPheValTYrrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyLys 660
 QY 1991 TACTTGGCGCTCAACACACCTGCGCTCCGCTACGCGCGGTGATGAGAGCGCGAG 2050
 |||||||
 Db 661 TYrrLeuArgLeuAsnThrLeuAlaSerLeuGlyTYrrAlaValAlaValIleAspLysArg 680
 QY 2051 GGCCTCTGTCAGCAGAGCGCTTCGTCAGAGGGCGCTGAAACCAAAATGGCGCAGGTG 2110
 |||||||
 Db 681 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnVal 700
 QY 2111 GAGATCGAGGACACAGGTGAGAGCGCTGCAATCTGCGCGGAGACTATGCTTATCAAC 2170
 |||||||
 Db 701 GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTYrrGlyPheIleAsp 720
 QY 2171 CTGAGCGGATGTCATGCGCTGCTACGCGCGCTTCCTGCTCATGGCGGCTA 2230
 |||||||
 Db 721 LeuSerArgValAlaAlaIleHisGlyTYrrPserTYrrGlyPheLeuSerLeuMetGlyLeu 740
 QY 2231 ATCCACAAAGCGCGGTTCAGGTGCGCATCGCGGGTGCCTCGCTACCGCTGTGATG 2290
 |||||||
 Db 741 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMet 760
 QY 2291 GCGTCACGACAGAGGTACACTGACGCGCTACATGACGCTCCCTGAGACACACGACGCGC 2350
 |||||||
 Db 761 AlAluYrrAspThrGlyTYrrThrGlnArgTYrrMetAspValProGluAsnAsnGlnHisGly 780
 QY 2351 TATGAGGGGGTTCGCTGCGCTGACGTGAGAGAGCTGCCATGAGCCCAACCGCGCTG 2410
 |||||||
 Db 781 TYrrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGlnProAsnArgLeu 800
 QY 2411 CTATCTCTCACAGCGCTTCGAGCAAAAGCTGCACTTTTCCACAAACTTCCTCGTC 2470
 |||||||
 Db 801 LeuIleLeuHISHisGlyPheLeuAspGluAsnValHisPhePheHISrHrAsnPheLeuVal 820
 QY 2471 TCCCAACTGATCCGAGCAGGAGAACTTACAGAGCTCCGATCTACCCCAAGAGAGAC 2530
 |||||||
 Db 821 SerGlnLeuIleArgAlaGlyLysProTYrrGlnLeuGlnIleTYrrProAsnGlnLysArgHis 840
 QY 2531 AGTATTCGCTGCTCGCGAGTTCGGGAGAGACATATGAGTACAGTTCCTCTACAG 2590

Db 841 SerIleArgCysProGluSerGlyGluHisTYrrGluValThrLeuLeuHISrPheLeuGln 860
 |||||||
 QY 2591 GAATACCTC 2599
 |||||||
 Db 861 GluTYrrLeu 863
 RESULT 2
 ABG61602
 ID ABG61602 standard; Protein: 892 AA.
 AC ABG61602:
 DT 12-AUG-2002 (first entry)
 DE Human DPP-2 splice variant #1.
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP-2;
 KW DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 OS Homo sapiens.
 PN W020031134-A2.
 PD 18-Apr-2002.
 PE 12-Oct-2001; 2001MO-US11874.
 PF 12-Oct-2000; 2000US-240117P.
 PR 12-Oct-2000; 2000US-240117P.
 PX (FERR) FERRING BV.
 PA Qi S, Aktinsanya KO, Riviere PJ, Junien J;
 PI WPI: 2002-444178/47.
 DR N-PSDB; ABK83333.
 DX
 DR
 DX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins; useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PS Disclosure; Page 76-78; 113pp; English.
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPPP proteins.
 XX
 SQ Sequence 892 AA;
 Alignment Scores:
 Pred. No.: 0
 Score: 4646.00 Length: 892
 Percent Similarity: 100.00% Matches: 863
 Best Local Similarity: 100.00% Conservative: 0
 Query Match: 96.89% Mismatches: 0
 DB: 23 Indels: 0
 Gaps: 0

US-09-976-674-4 (1-2617) x ABG61602 (1-892)

QY 11 ATGGCCACGACCGGAGCCCAAGCGCCGAGCGAGCGAGCGCGGACAGATGACCG 70
Db 30 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaAlaThrAspAspPro 49
QY 71 GCGCGCCGCTTCCAGGTGACGAAGACCTGTGGAGCGGGCTCCGAGCATCATCACGGC 130
Db 50 AlaAlaArgPheGlnValGlnLysHisSerThrAspArgLeuAlaGserLleIleHisGly 69
QY 131 AGCGGCAAGTACTGGGCTTCATTTGTCAACAGGCGGCCCAAGCATTCACGTTTGTCAG 190
Db 70 SerArgLysThrSerGlyLeuIleValAlaSnLysAlaProHisAspPheGlnPheValGln 89
QY 191 AAGCGATGAGTGTGGGCGCCCACTCCACCGGCTCTACTACCTGAGGGAAGGCCATATGGC 250
Db 90 LysThrAspLysSerGlyProHisSerHisArgLeuThrTyrLeuGlyMetProTyrGly 109
QY 251 AGCGGAGAACTCCCTCTCTACTGTAGATTCAGATTCACGAAGAGTCCGGAAAGAGCTGTG 310
Db 110 SerArgGlnAsnSerLeuLeuThrSerGluLeuProLysValArgLysGluAlaLeu 129
QY 311 CTGTCTCTGTCTGGAGACAGATGCTGATCATTTCCAGGCGCACGCCCCACCATGGGGTC 370
Db 130 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisLysGlyVal 149
QY 371 TACTCTGGAGAGAGAGACTGTGAGGAGCGGAACGCGCTGGGGGTCTTCGCGCATCAC 430
Db 150 TyrSerArgGlnGluGlnLeuLeuArgGluArgLysArgLeuGlyValPheGlyLleThr 169
QY 431 TCCTAGACTTCCACAGCGAGAGTGGCCCTCTCTCTCCAGGCGACCAACAGCTCTTTC 490
Db 170 SerTrpAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
QY 491 CACTGTGGCAGCGCGGCGCAGAAAGGCTTCATGCTGCCCTAAGAACCGCTGGAAATC 550
Db 190 HisLysArgAspArgLysGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 209
QY 551 AAGACGAGTGCATCAGGCGCCCGGATGAGCCCAAAATCTGCCCTCGCGGACCTGCTTC 610
Db 210 LysThrGlnLysSerGlyProArgMetAspProLysLleCysProAlaAspProAlaPhe 229
QY 611 TTCTCTTCATCAATAACAGCGACTGTGGTGGCCCAACATCGACAGCGGAGGAGCGG 670
Db 230 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlnGluArg 249
QY 671 CGGTCGACTTTCGCGCCCAAGGTTTATCCATGCTCCTGATGATGCCCAAGTCTGGGGT 730
Db 250 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 269
QY 731 GTGGCCACTTCGTATACAGAGAGATTCGACCGCTTCACCTGGTACTGGTGGTGGCC 790
Db 270 ValAlaThrPheValIleGlnGlnGluPheAspArgPheThrGlyTyrTrpTrpCysPro 289
QY 791 ACAGCCTCTGGGAAGGTTCAAGAGGCTCAAGAGCTGCGAATCTGTATGAGGAATC 850
Db 290 ThrAlaSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnGluVal 309
QY 851 GATGAGTCCAGGTGAGGTGATCATTCAGTCCCTCTCTGCTGCTGCTGCTGCTGCTGCTG 910
Db 310 AspLysSerGlnValGlnValIleHisValProSerProAlaLeuGlnGluArgLysThr 329
QY 911 GACTCGATGGGTACCCAGAGACAGGACAGAAATCCAAAGATTTGGCTTGAATCTGGCT 970
Db 330 AspSerTrpArgLysTrpArgLysThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
QY 971 GAGTTCAGACTGACACGACGAGAGATGCTTTCGACCCAGAGAAAGAGAGCTGGTCCAG 1030
Db 350 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnLysGlnLeuValGln 369
QY 1031 CCTTTCAGCTCGCTGTTCCGGAAGGTGAGTACATGCCGAGGCGGGGTGACCCGGGAT 1090
|||||

Db 370 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGluTyrTrpArgAsp 389
QY 1091 GGCAAATATACGCTTGGGCTATGTTCTGTGACGGGCGCCAGCAGATGGCTCTGCTCTC 1150
Db 390 GlyLysTyrAlaThrAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 409
QY 1151 CTCGCCCGGCGCTGTTCATCCGAGACAGAAATGAGAGCAGCGGCTAGCCTTGCC 1210
Db 410 LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGlnArgLeuAlaSerAla 429
QY 1211 AGAGCTCTCCCGAGATGTCCAGCCCTATGTGTGACAGAGAGCTCACCACGCTGCG 1270
Db 430 ArgAlaValProArgAsnValGlnProTyrValValTyrGlnGluValIleThrAsnValThr 449
QY 1271 ATCATGTTCATGACATCTTATCCTTCCCTCCCAATCAGAGGAGAGAGAGCTTGC 1330
Db 450 IleAsnValHisAspIlePheTyrProPheProGlnSerGlnGlyLysArgGlnLeuCys 469
QY 1331 TTTCTCGCGCCCAATGATGATCAAGACCGGCTTGTGCCATTTGTACAAAGTACCGCGTT 1390
Db 470 PheLeuArgAlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValIleThrAlaVal 489
QY 1391 TTTAAATCCAGGCTACGATGAGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1450
Db 490 LeuLysSerGlnGlyTyrAspTrpSerGlnProPheSerProGlnLysGlnAspGluPheLys 509
QY 1451 TGCCCCATTAAGAGAGATGTGCTGTACACGCGGTGAATGGAGGAGGAGGAGGAGGAG 1510
Db 510 CysProIleLysGlnGluIleAlaLeuThrSerGlyLysTrpGlnValLeuAlaArgHis 529
QY 1511 GGCTCCAAATCTGGGTCAATGAGAGACAAAGCTGTGACTTCCAGGCGACCAAGGAC 1570
Db 530 GlySerLysIleTrpValAsnGlnGluThrLysLeuValTyrPheGlnGlnTyrLysAsp 549
QY 1571 ACGCGCTGGAGACGACGCTCTACGTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 1630
Db 550 ThrProLeuGlnHisLysLeuTyrValValSerTyrGlnAlaAlaGlyGluIleValArg 569
QY 1631 CTCACACGCGCGGCTTCTCCATAGCTGTGCTCATGAGCCAGAACTTGCATGTTGCTC 1690
Db 570 LeuThrThrProGlnPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 589
QY 1691 AGCCACTACAGACGCTGAGACGCGCGCTGCTGACAGCTTACAAAGCTGAGCGGCGCC 1750
Db 590 SerHisTyrSerSerValSerThrProCysValHisValTyrLysLeuSerGlyPro 609
QY 1751 GAGCAGACCGCCGTCAGACAGAGCGCGCTTGGGCTGAGATGAGGAGGAGCGGACG 1810
Db 610 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaAlaSer 629
QY 1811 TGCCCCCGGATTATGTTCTCCAGAGATCTTCCATTTCCACAGCGCTGAGATGCGG 1870
Db 630 CysProProAspArgValProProGlnIlePheHisPheHisThrArgSerAspValArg 649
QY 1871 CTCTAGGCGATGATCTCAAGCCCGACGCTTGCAGCGAGGAGAAAGCAAGCCAGCTC 1930
Db 650 LeuTyrGlnMetIleTyrLysProHisAlaLeuGlnProGlnLysLysHisProThrVal 669
QY 1931 CTCTTGTATATGAGGCGCCCGAGGCTGACGTGTAATTAATCTTCAAGGATCAAG 1990
Db 670 LeuPheValIleTyrGlnGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyLleLys 689
QY 1991 TACTTGGCGCTCAACACATGCGGCTCCCTGGGCTACGCGCTGTGTGATGAGCGGAG 2050
Db 690 TyrLeuArgLysAsnTrpLeuAlaSerLeuGlnTyrTrpAlaValValIleAspGluArg 709
QY 2051 GCGTCTGTACGAGGCGCTTGGTTCGAAGGGGCTGAAGAAACAAATGGGCGAGGTG 2110
Db 710 GlySerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnVal 729
QY 2111 GAAATCAGAGACAGAGTGGAGGCGCTGACATGCTGTGGCCGAGAAATTTGGCTTATCGAC 2170
Db 730 GluIleGlnAspGlnValGlnGlnGluGlnPheValAlaGlnLysTyrGlnPheIleAsp 749
|||||

QY 2171 CTGAGCCAGTTGCGATCATGCGTGGGCTTCAGCGGGGCTTCCTGCGTCATGGGGCTA 2230
 |||||||
 Db 750 LeuSerArgAlaAlaIleHisGlyTrpSerTrpGlyLeuLeuSerLeuMetGlyLeu 769
 QY 2231 ATCCACAGCCCGGAGTGTTCAGGTGGCCATCCGCGGGTCCCGGTCACCGTGTGATG 2290
 |||||||
 Db 770 IleHisLysProGlnValPheLysValAlaIleLeuGlyAlaProValThrValTrpMet 789
 QY 2291 GCGTACGACAGCGGTACACTGACGCGTACATGAGCGTCCCTAGAACACACGACGCGC 2350
 |||||||
 Db 790 AlaTrpAspTrpGlyTrpGlyTrpGlyTrpGlyTrpGlyTrpGlyTrpGlyTrpGlyTrp 809
 QY 2351 TATAGAGCGGCTCCGCGCGCTGACGTCGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 2410
 |||||||
 Db 810 TrpGlnAlaGlySerValAlaLeuHisValGlnLysLeuProAsnGlnProAsnArgLeu 829
 QY 2411 CTGATCCCTCCAGCGCTTCCTGAGCAAAACGTGACATTTTCCACAACTTCCTGCTC 2470
 |||||||
 Db 830 LeuIleLeuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuVal 849
 QY 2471 TCCCACTGATCGACGAGGAAACCTTACAGCTCCAGATCTACCCGACGAGACAC 2530
 |||||||
 Db 850 SerGlnLeuIleArgAlaGlyLysProTrpGlnLeuGlnIleTrpProAsnGlnArgHis 869
 QY 2531 AGTATTCGCTCCCGGAGTCCGCGGAGACACTATGACGTCAGCTTCCTGCTGACG 2590
 |||||||
 Db 870 SerIleArgLysProGlnSerGlyLysHisTrpGlnValThrLeuLeuHisPheLeuGln 889
 QY 2591 GAATACCTC 2599
 |||||||
 Db 890 GluTrpLeu 892
 RESULT 3
 ABG61604
 ID ABG61604 standard; Protein: 892 AA.
 XX
 AC ABG61604:
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-2 splice variant #3.
 XX
 KW Human: serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KW DPRP; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 KW
 OS Homo sapiens.
 XX
 PN W020023134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI: 2002-444178/47.
 XX
 DR N-PSDB; ABK83335.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 XX

PS Disclosure; Page 81-84; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPRP)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
 XX
 SQ Sequence 892 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 892
 Score: 4646.00 Matches: 863
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 96.89% Indels: 0
 DB: 23 Gaps: 0
 US-09-976-674-4 (1-2617) x ABG61604 (1-892)
 QY 11 ATGGCCACACCGGACCCCAAGCGCGACGAGCGAGCGCGCGCGCGCGCGCGCGCG 70
 |||||||
 Db 30 MetAlaThrTrpGlyThrProThrAlaAspArgGlyAspAlaIleAlaThrAspAspPro 49
 QY 71 GCCGCCGCTTCCAGGTGACAGACACTGTGGAGCGGCTCCGAGCATATCCAGCGC 130
 |||||||
 Db 50 AlaIleArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 69
 QY 131 AGCCGCAAGTACTCGGCGCTCATGTGTCAACAGCGCGCGCGCGCGCGCGCGCGCG 190
 |||||||
 Db 70 SerArgLysTrpSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 89
 QY 191 AAGACGATGAGTGTGGCGCCCACTCCACCGCGCTCTACTAGCTGGAGATGCCATATG 250
 |||||||
 Db 90 LysThrAspGlnSerGlyProHisSerHisArgLeuTrpTrpLeuGlyMetProTrpGly 109
 QY 251 AGCCGAGAACTCCCTCTCTACTGTGAGATTCGCAAGAGCGCGCGCGCGCGCGCG 310
 |||||||
 Db 110 SerArgGlnAsnSerLeuLeuTrpSerGlyLeuProLysValArgLysGlnAlaLeu 129
 QY 311 CTGCTCCCTGCTGGAAGCAATGCTGATATTCGAGCGCGCGCGCGCGCGCGCGCG 370
 |||||||
 Db 130 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 149
 QY 371 TACTCTGGAGAGGAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 430
 |||||||
 Db 150 TrpSerArgGlnGlnGlnLeuLeuArgGlnAlaGlyLysAlaGlnGlyValPheGlyIleThr 169
 QY 431 TCTAGCACTTCCACAGGAGAGTGGCTTCTCTTCCAGGCGCGAGACAGCGCTTTC 490
 |||||||
 Db 170 SerTrpAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
 QY 491 CACTGTCCGACGCGCGGAGAGAGCGGCTTCATGGTGTCCCTATGAAACCGCTGGAAT 550
 |||||||
 Db 190 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 209
 QY 551 AAGACCCAGTGTGAGCGCGCGGATGAGCCCAAAATCTGCGCTGCGAGCGCTGCTTC 610
 |||||||
 Db 210 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
 QY 611 TTCTCTTCATCATATACAGCGACCTGTGGGTGGCCCAATGACGAGCGGAGCGG 670
 |||||||
 Db 230 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGlnThrGlyGlnGlnArg 249

QY 671 CGGTGACCTTCTGCACCAAGTTTATCCATGCTCGATGACCCCAAGCTGCGGCT 730
DB 250 ArgLeuThrPheCysSHSgInGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 269
QY 731 GTGGCCACCTTCGTATACAGAGAGAGTTCAGCGCTTCACTGGGTACTGGTGGTGGCC 790
DB 270 ValAlaIaThrPheValIleGInGInGInPheAspArgPheThrGlyTyrTrpPheCysPro 289
QY 791 ACAGCCCTCTGGGAAGTTACAGAGGCTCAAGAGCTGGCAATCTGTATGAGAAATC 850
DB 290 ThrAlaSerTrpGInGlySerGInGlyLeuLysThrLeuArgIleLeuTyrGInGInVal 309
QY 851 GATAGAGTCCAGGTGAGAGGTATTCAGTCCCTCTCTGGGCTAGAGAAAGAAAGACG 910
DB 310 AspIuSerGInValGInValIleHisValProSerProAlaLeuGInGInValArgLysThr 329
QY 911 GACTCGATCGGTACCCAGAGACAGGAGCAAGATCCCAAGATGGCTTGAATCTGGCT 970
DB 330 AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
QY 971 GAGTTCAGACTGACAGCCAGGCGAAGATCGTCTGACCCAGAGAAAGAGAGCTGTGACG 1030
DB 350 GluPheGInThrAspSerGInGlyLysIleValSerThrGInGInLysGInLeuValGIn 369
QY 1031 CCTTCAGCTCGCTGTTCCCGAAGGTGAGTACATCCCGAGGCGCGGTGACCCGGGAT 1090
DB 370 ProPheSerSerLeuPheProLysValGInLutyrIleAlaArgAlaGlyTrpThrArgAsp 389
QY 1091 GGCAGAAATACGCTGGGCGATGTTCTTGACCGGCGCCAGAGTGGTCCAGCTGTCTC 1150
DB 390 GInLysTyrAlaTrpAlaMetPheLeuAspArgProGInGInTrpLeuGInLeuValLeu 409
QY 1151 CTCGCCCGGCGCTTTCATCCCGAGACAGAGATGAGAGAGCGGCTAGCCCTTGCC 1210
DB 410 LeuProProAlaLeuPheIleProSerThrGInAsnGInGInValArgLeuAlaSerAla 429
QY 1211 AGAGCTTCCCGAGAGATGTCACGCCGTATGTGGTGTACAGAGAGGTGACCACGCTGG 1270
DB 430 ArgGInAlaValProArgAsnValGInProTyrAlaValTyrGInGInValIleThrAsnValTrp 449
QY 1271 ATCAATGTTATGACATCTTATCCCTTCCATCCCGCAATCAGAGGAGAGAGAGCTGTGC 1330
DB 450 IleAsnValHisAspIlePheTyrProPheProGInSerGInGInLysAspGInLeuCys 469
QY 1331 TTTCTCGCGCCATGATGATGACAGACGCGCTTGCATTGTGACAAAGTCACCGCGCTT 1390
DB 470 PheLeuArgAlaAsnGInCysLysThrGlyPheCysHisLeuTyrLysValIleThrAlaVal 489
QY 1391 TTTAAATCCCAAGGCTTCAGTGGAGTGAGACCTTCAGCCCGGAGGAAAGATTTAAG 1450
DB 490 LeuLysSerGInGlyTyrAspTrpSerGInProPheSerProGInGInLysAspGInPheLys 509
QY 1451 TGCCCCCTTAAAGAGAGATGCTGTGACAGAGCGGTAAATGGAGGTTTGGCAGAGCAC 1510
DB 510 CysProIleLysGInGInIleAlaLeuThrSerGInLutTrpGInValIleAlaIaArgHis 529
QY 1511 GGCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGATCTCCAGGCGACCAAGAGAC 1570
DB 530 GlySerLysIleTrpValAsnGInGInLutThrLysLeuValTyrPheGInGInLutThrLysAsp 549
QY 1571 ACCGCGGTGAGAGACCACTCTACGTGGTACATATAGAGCGCGCGCGAGATCTGACGC 1630
DB 550 ThrProLeuGInHisIleValTyrValIaSerTyrGInLutAlaIaGInGInLutLeuValArg 569
QY 1631 CTCACACGCGCGCTTCTCCATAGCTGTGCTCATGAGCAGAACTCGCATGTTGCTC 1690
DB 570 LeuThrThrProGInPheSerHisSerCysSerMetSerGInAsnProPheAspMetPheVal 589
QY 1691 AGCCACTACAGACAGCGTACAGCGCGCGCTGTGTCACAGCTTACAGAGCGCGCCC 1750
DB 590 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGInPro 609
QY 1751 GACGAGACCGCCCTGCACAAAGACGCGCGCTTCTGGGCTGACATGATGAGAGCGCCAGC 1810

DB 610 AspAspAspProLeuHisLysGInProArgPheTrpAlaSerMetMetGInAlaAlaSer 629
QY 1811 TGCCCCCGGATTTATGTTCTCCAGAGATCTTCATTTCCACAGCGCGCTGGATGGCGG 1870
DB 630 CysProProAspTyrValIleProProGInIlePheHisIleHisThrArgSerAspValArg 649
QY 1871 CTCTACGGCATGATCTAACAGCCCGCCCTTCAGCGCGAGGAGAAAGAACACCCACCGTCC 1930
DB 650 LeuTyrGInMetIleTyrLysProHisAlaLeuGInProGInLysLysHisProThrVal 669
QY 1931 CTCTTTTATATGAGAGCGCCCGAGAGTCCAGCTGGTGAATATCTCTTCAAGGATTCAG 1990
DB 670 LeuPheValTyrGInGlyTyrProGInValGInLeuValAsnAsnSerPheLysGInLys 689
QY 1991 TACTTCGGGCTCAACACACATCGGCTCCCTGGGCTACCGCGCTGGTGTGATTCAGAGCAG 2050
DB 690 TyrLeuArgLeuAsnThrLeuAlaSerLeuGInTyrTrpAlaValValIleAspGInArg 709
QY 2051 GGCTCTGTACGAGGCGCTTCGGTTGGAAGGGCCCTGAAGAAACCAATGGCGCAGGTG 2110
DB 710 GlySerCysGInArgGInLeuArgPheGInGInLutAlaLeuLysAsnGInMetGInGInVal 729
QY 2111 GAGATCGAGACCGAGGTGAGAGGCGCTGACATGCTGGCGCGAGAAATGTGGCTTCATCGAC 2170
DB 730 GluIleGInAspGInValGInGInLysGInPheValAlaGInLysTyrGInPheIleAsp 749
QY 2171 CTGACCGGATGGCATCATGCTGCTGCTACGAGGCGCTTCTCTGCTGATGAGGCGCTA 2230
DB 750 LeuSerArgValAlaIleHisGInLutTrpSerTyrGInLysPheLeuSerLysMetGInLeu 769
QY 2231 ATCCACAAGCCCGAGTGTTCAGAGTGGCCATCGCGGGTCCCGGCTACCGCTGTGGATG 2290
DB 770 IleHisLysProGInValIlePheLysValAlaIleAlaGInAlaProValIleThrValTrpMet 789
QY 2291 GCTTACGACACAGGCTACATGAGCGGTACATGAGAGCTCCCTGAGAACACACAGCAGCGC 2350
DB 790 AlaTyrAspPheArgTyrTrpGInArgTyrMetAspValProGInAsnAsnGInHisGly 809
QY 2351 TATGAGCGGCGTCCGTTCCGTCGACAGTGGAGAAAGCTGCCAATGAGCCCAACCGGCTTG 2410
DB 810 TyrGInAlaIleLysValAlaLeuHisValGInLysLeuProAsnGInLutProAsnArgLeu 829
QY 2411 CTATCTCCACGCGCTTCGTGACGAAACGTGCACATTTTTCACACAAATCTCTGCTC 2470
DB 2471 TCCCAACTGATCCGAGAGGAAACCTTACCAGCTCCAGATCTTACCCCAAGAGAGACAC 2530
QY 2531 AGTATTCCGTGCGCCCGAGTGGCGGAGCACTATGAAGTCACTTGGTGCATTTACAG 2590
DB 870 SerIleArgCysProGInSerGInGInHisTyrGInValIleThrLeuLeuHisPheLeuGIn 889
QY 2591 GAATACCTC 2599
DB 890 GluTyrLeu 892
RESULT 4
ID AAE24168 standard; Protein; 969 AA.
AC AAE24168;
XX
DT 23-SEP-2002 (first entry)
XX Human dipeptidyl peptidase 9 (DPP9) protein.
XX Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme.

XX	Homo sapiens.
OS	
XX	
FH	Key
FF	Misc-difference 374
FT	/note= "Encoded by GAA"
XX	
PN	WO200234900-A1.
XX	
PD	02-MAY-2002.
XX	
PF	29-OCT-2001; 2001WO-AU01388.
XX	
PR	27-OCT-2000; 2000AU-0001078.
XX	
PA	(UNSY) UNIV SYDNEY.
XX	
PI	Abbott CA, Gorrell MD;
XX	
DR	WPI: 2002-454646/48.
N-PSDB:	AAD38954.
XX	
PT	New dipeptidyl peptidase (DPP) peptides, useful for screening
PR	inhibitors of DPP catalytic activity, which may be employed to treat
PT	e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PR	rejection and HIV infection -
XX	
PS	Claim 1; Fig 4; 91pp; English.
XX	
CC	The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC	polynucleotides encoding such proteins. The DPP peptides are useful for
CC	screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC	for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC	rejection and HIV (human immuno deficiency virus) infection. The present
CC	sequence is human DPP9 protein.
XX	
SQ	Sequence 969 AA:
Alignment Scores:	
Pred. No.:	0 Length: 969
Score:	4636.00 Matches: 861
Percent Similarity:	99.88% Conservative: 1
Best Local Similarity:	99.77% Mismatches: 1
Query Match:	96.68% Indels: 0
DB:	Gaps: 0
US-09-976-674-4 (1-2617) x AAE24168 (1-969)	
QY	11 ATGGCCACCAACCGGAGCCCCAAGCGGCCGAGGCAGCGCAGCCGCACAGATGACCCG 70
Dd	
Db	107 MetAlaThrThrfcIlyTrfProtnrAlaaSPAr9gLyAspAlaAlaAlaThraSPaSPro 126
QY	71 GCCGCCGGTTCAGAGGTGCAGAAGCACTCGTGAGCGGCTCCGAGCATCATCCAGCGC 130
Dd	
Db	127 AlaAlaArgPheInvalnIglnLysHisSerTriPaSpolyLeuAtySerIlelleHisSely 146
QY	131 AGCCGGAAGTAAGTCTGGGCGCTCATTTGTCAACAAGCGCCGCCACGACTTCACATTGTGCAG 190
Dd	
Db	147 SerArjLysTyrsErgIyleuIlleValaAsnLysLaPrhoHisAspPheGlnPhVaIglIn 166
QY	191 AAGACGAGTAGTCTGGGGCCCACTCCACCGGCTTACTACTGCTGGGGAATGCCATATGAGC 250
Dd	
Db	167 LysThraSpolusertglYPrOHIsSerHisArgdeutyrtYrLeuGlYMeIProttyclY 186
QY	251 AGCCGAGAGACTCCCTCTCTACTCTGAGATTGCCAAGAAAGTCCGGAAGAGGCTCTG 310
Dd	
Db	187 SerArjGluAnsSerleuLeutyrSerGluiLeProlysValaArgLyGluAlaLeu 206
QY	311 CTGCTCTCTGTCGCGAAAGAGATGCTGGATCATTTCCAGGCGAGGCCACCATGGGGTGC 370
Dd	
Db	207 LeuLeuLeuSertrPyssglImeLeuAspHisPheGlnAlatrnProHisHisGlyVal 226
QY	371 TACTCTGGGAGGAGAGCACTGTGAGGAGCGGAACGCTGGGGGCTTTCGGCATCACCC 430

[illegible]

Db 587 CysProIleIysGluIleAlaLeuThSerSergIuTrpGluValLeuAlaArgHis 606
 QY 1511 GGCATCAAGATCGTGGTCAATAGAGACAGCAAGCTGTGTACTTCCAGGCAAGAC 1570
 Db 607 GlySerIuSileTrpValAsnGluGluThrIuSleuValIuTrpHeGluGlyThrIuSasp 626
 QY 1571 AGCGCGGTGAGACACACCTCTACGTGTGACATATGAGGGCGCGGAGATCGTACGC 1630
 Db 627 ThrProIeuGluHisIleuTyTrValValSerIuGluAlaIaGluIleValArg 646
 QY 1631 CTCACACAGCGCGCTCTCCATAGCTGCCTCCATGAGCCGAACCTCGACATGTGTCTC 1690
 Db 647 LeuThrTrpProGluTrpHeSerHisSerIuSerMetSerGluAsnPhaSerMetPheVal 666
 QY 1691 AGCCACTACACAGCGGTGAGACAGCGCGCTGTGACATCTACAGCTACAGCGCGCC 1750
 Db 667 SerHisIuSerSerValSerThrProProCysValHisValIuTrpIuSleuSergIuPro 686
 QY 1751 GAGACGACACCCCTGACACAGACAGCGCGCTGTGCGCTACATGATGAGAGCAGCAGC 1810
 Db 687 AspAspAspProIeuHisIuSglInProArgPheTrpAlaSerMetGluAlaAlaSer 706
 QY 1811 TGCCCCCGGATTATGTCTCCAGAGATCTTCATTTCCACAGCGCGCTCGGATGTGGG 1870
 Db 707 CysProProAspIuTrpValProProGluIlePheHisPheHisTrpArgSerAspValArg 726
 QY 1871 CTCATACGCGATGATCTACAGACCCACCGCTTGACAGCGAGGAGAAAGCACCCACGCTC 1930
 Db 727 LeuIuTrpGluMetIleTyTrpIuSglProHisAlaIeuGluInProGluIuSgluHisIuProThrVal 746
 QY 1931 CTCCTTTATATGAGAGCGCGCCAGGTGACAGCTGTGAATACTCTTCMAAGCATCAAG 1990
 Db 747 LeuPheValIuTrpGluIuSglProGluAlaGluIleuValaAsnAsnSerPheIuSglIuLeuS 766
 QY 1991 TACTTGGCGCTCACACACACCTGGCTCCCTGGGTACAGCGCGGTGTGATGAGGCGCAGG 2050
 Db 767 TyTrpLeuArgIeuAsnThrLeuAlaSerIeuGluIuTrpAlaValaIleIuAspIuArg 786
 QY 2051 GGCCTCTGTACGAGAGGCTTCGCTTCGAAAGGCGCCCTGAAACCAAAATGGCGCAGGTG 2110
 Db 787 GlySerCysGlnArgGluLeuArgPheGluIuGluAlaIeuIuSgluAsnGluMetGluVal 806
 QY 2111 GAGATCGAGGACCAAGGTGAGAGGCGCTGACATTCGTGCGCGCAAGATGCTTCATCGAC 2170
 Db 807 GluIleGluAspGlnValaGluGluIleuGluInPheValaIaGluIuSgluIuPheIleAsp 826
 QY 2171 CTGACCGAGTGGCATCATGAGGTGGTCTACAGGGGCGCTCTCTGCTCATGGGGCTA 2230
 Db 827 LeuSerArgValaAlaIleHisGluIuTrpSerIuTrpGluIuPheLeuSerIeuMetGluLeu 846
 QY 2231 ATCCACAAGCCCCAGGTGTTCAAGGTGGCCATTCGCGGGTCCCGGTCCACCCGTCTGATG 2290
 Db 847 IleHisIuSglProGluInValaPheIuSglValaAlaIleAlaGluAlaProValThrValTrpMet 866
 QY 2291 GCCATACGACACAGGCTACACTGAGCGCTACATGGAAGTCCCTGAGAACACACAGCAGCGC 2350
 Db 867 AlaTyArgSprIuTrpGluIuTrpGluArgTyMetAspValProGluAsnAsnGluHisGlu 886
 QY 2351 TATAGAGGGGCTTCGCGGCGCTCGACGTGGAGAAAGTGGCCATGAGGCCAACCGGCTTG 2410
 Db 887 TyTrpIuAlaGluSerValaAlaLeuHisValaGluIuSgluLeuProAsnGluProAsnArgIeu 906
 QY 2411 CTATTCCTCCAGGCTCTCTGAGCAAAACGTGCACCTTTTCCACACAAACTCTCTGCTC 2470
 Db 907 LeuIleLeuHisGluPheLeuAspGluAsnValaHisPhePheHisTrpAsnPheLeuVal 926
 QY 2471 TCCCACTGATCGAGCAGGAAACCTTACAGCTCCAGATCTACCCCAAGACAGACAC 2530
 Db 927 SerGlnLeuIleArgAlaGluIuSgluProTyGlnIleuGlnIleTyTrpCAsnGluArgHis 946
 QY 2531 AGTATTCGCTGCGCGGAGTGGGAGAGCATATGAAAGCAGTGGCTCCATTTCTACAG 2590
 Db 947 SerIleArgCysProGluSerGluIuHisIuTrpGluIuValaThrLeuLeuHisPheIeuGln 966

QY 2591 GAATACCTC 2599
 Db 967 GluTyIleu 969
 RESULT 5
 ID ABG61607 standard; Protein: 879 AA.
 XX
 AC ABG61607;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPPR-2 splice variant #6.
 DE
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR;
 KW DPPR; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinnesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI: 2002-444178/47.
 DR N-PSDB: ABK83338.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins; useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 PS Disclosure; Page 91-93; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypertension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinasias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPPP proteins.
 CC
 XX
 SQ Sequence 879 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 879
 Score: 4558.50 Matches: 850
 Percent Similarity: 98.49% Conservative: 0
 Best Local Similarity: 98.49% Mismatches: 0
 Query Match: 95.07% Indels: 13
 DB: 23 Gaps: 1

US-09-976-674-4 (1-2617) x ABG61607 (1-879)

QY	11	ATGSCCACCACGGGAGCCCAACGCGCCGACGAGCAGCCGCCACAGATGACCG	70
Db	30	MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaAlaThrAspAsp	49
QY	71	GGCGCGCGCTCCAGGGCAGAAAGCACTGCTGGGAGCGGGCTCCGGAGATCATCCAGGC	130
Db	50	AlaAlaArgPheGlnValGlnLysHisSerThrPaspGlyLeuArgSerIleHisGly	69
QY	131	AGCGCAGACTACGCGGCCCATTTGTCACAAAGGCGCCCAACGATTCACATTTGTGCAG	190
Db	70	SerArgLysIyrSerGlyLeuIleValasnLysAlaProHisAspPheGlnPheValGln	89
QY	191	AAGACGATAGTCTGGGCCCCCACTCCACGCGCTCTACTACCTGGGAATGCCATATGC	250
Db	90	LysThrAspGlnSerGlyProHisSerHisArgLeuIyrTYrLeuGlyMetProTYrGly	109
QY	251	AGCGGAGAAACTCCCTCTACTCTGTGAGATTTCCCAAGAGTCCGGAAGAGCTCTG	310
Db	110	SerArgLysasnSerLeuLeuIyrSerGluIleProLysValArgLysGluAlaLeu	129
QY	311	CTGCTCTGTCGTCGAAGCAGATGCTGGATTCATTCCAGAGCCAGCCGCCACCATGGGGTC	370
Db	130	LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal	149
QY	371	TACTCTCGGAGAGGAGACGTCTGAGGAGACGGAAACGCTTGGGGGTCTTGCGATCAC	430
Db	150	TyrSerArgGlnGlnGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr	169
QY	431	TTCCACGACTTCCACAGCGAGATGGCGCTTCTCTCTCCAGGCGCAGAACAGCTCTTC	490
Db	170	SerIyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe	189
QY	491	CACGTGCGCGCGCGCGCAGAACGGCTTCATGAGTCCCTCATGAACCCCTGGAATC	550
Db	190	HisIyrAspArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle	209
QY	551	AAGACCCAGTGCCTCAGGCCCCCGGATGGAACCCCAAAATCTGGCCCTGCCAGCCTGCCTC	610
Db	210	LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe	229
QY	611	TTTCTCTTCAATCAATAACACGCACTGTGGGTGGCCACATGGAGACAGCGAGAGCGG	670
Db	230	PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg	249
QY	671	CGGTCGACCTTCTGCGCAACGAAGTTTATTCATATGCTCGGAGACCCCAAGTCGGCGGT	730
Db	250	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerIleAlly	269
QY	731	GTGGCCACCTTCGCATACAGAAAGATTCTGACCGCTTACAGGGTACTGGTGTCGCC	790
Db	270	ValAlaThrPheValIleGlnGlnGluPheAspArgPheThrGlyTYrTrpTrpCysPro	289
QY	791	ACAGCTCTCGGAAGGTTCAGAGGGCTTCACAGCCTGCGAATCTGTATGAGAACTC	850
Db	290	ThrIaSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTYrGluVal	309
QY	851	GATAGTCCTCGAGGGAGAGTCATTCACAGTCCCTCTCCCTGCGCTAGAAAAGAAAGACG	910
Db	310	AspGlnSerGlnValGlnValIleHisValProSerProAlaLeuGluGluThrGlyThr	329
QY	911	GACTCTGATCGGTACCCACAGACAGCAGCAAGAAATCCCAAGATTGCTTGAATGGCT	970
Db	330	AspSerIyrArgIyrProAlaGlnThrGlnGlySerLysAsnProLysIleAlaLeuLysLeuAla	349
QY	971	GAGTTCAGACTGACACCGAGGGCAAGATGCTCTGCACCCAGAGAGAGAGCTGGTGCAG	1030
Db	350	GlnPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluTYrGluValGln	369
QY	1031	CCCTTCAGACTGCTGCTTCCCGAAGGTGAGATCATGCGCAGGCGCGGTGACCCGGGAT	1090
Db	370	ProPheSerSerLeuPheProLysValGluTYrIleAlaArgAlaGlyTrpThrArgAsp	389

[illegible]

```
|||||
Db 737 LeuSerArgValAlaIleHisGlyTyrPseTyrGlyGlyPheLeuSerLeuMetGlyLeu 756
QY 2231 ATCCACAGCCCGGAGGTCCTCAGGTGGCGATGCGGGTCCCGGTCACCGTCTGGATG 2230
Db 757 ILeHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValAlaThrValTyrMet 776
QY 2291 GCCTACGACACAGGCTACCTAGCGCTACATGGACGCTCCCTGAGAACACAGCAGCGC 2350
Db 777 AlaTyrAspThrGlyTyrThrGlnArgTyrMetAspValAlaProGlnAsnGlnHisGly 796
QY 2351 TATGAGCGGGGTTCCGTCGGCCCTGCACAGTGGAGAGAGTCCCAATGAGCCCAACCGCTTG 2410
Db 797 TyGlnAlaGlySerValAlaLeuHisValGlyLysLeuProAsnGlnProAsnArgLeu 816
QY 2411 CTATATCCCTCAGCGCTCTCGAGCAAAAGCTGATTTTCCACAAATTCCTGTC 2470
Db 817 LeuIleLeuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuVal 836
QY 2471 TCCCACTGATCGAGCAGGAGAAACCTTACACAGCTCCAGATCTACCCCAACAGAGACAC 2530
Db 837 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnArgHis 856
QY 2531 AGATTCGGCTGCCCGGAGTGGCGAGCAGCTATGAACTCACGTTGCTGCACTTTCTACAG 2590
Db 857 SerIleArgCysProGlnSerGlyGlnHisTyrGlyValThrIleLeuHisPheLeuGln 876
QY 2591 GAATACCTC 2599
Db 877 GluTyrIleu 879

RESULT 6
ABG61608
ID ABG61608 standard; Protein; 879 AA.
XX
AC ABG61608;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-2 splice variant #7.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychiatric disorder; neurological disorder;
KW dyskinnesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PE 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
PI O1 S, Akinsanya KO, Riviere PJ, Junten J;
DR WPI; 2002-444178/47.
DR N-PSDB; ABR83339.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
PS Disclosure; Page 95-96; 113pp; English.
XX
```

```
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC biliary, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinasias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 879 AA:
Alignment Scores:
Pred. No.: 0 Length: 879
Score: 4558.50 Matches: 850
Percent Similarity: 98.49% Conservative: 0
Best Local Similarity: 98.49% Mismatches: 0
Query Match: 95.07% Indels: 13
DB: 23 Gaps: 1

US-09-976-674-4 (1-2617) x ABG61608 (1-879)
QY 11 ATGGCCACACCGGGAGCCCGACGCGGCGAGGCGAGCGAGCGCCAGATCCGCG 70
Db 30 MetAlaThrThrGlnThrGlnProThrAlaAspArgGlyAspAlaAlaAlaThrAspPro 49
QY 71 GCCGCCCGCTTCCAGGTGCAGAGACACTGCTGGAGCGGCTCCGAGCATCATCCAGCGC 130
Db 50 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 69
QY 131 AGCCGCAAGTACTCGGGCTTCATTTGTACACAGGCGCCCGACGACTTCCAGTTTGTCAG 190
Db 70 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 89
QY 191 AAGACGATGAGTCTGGGCGCCCACTCCACCGGCTCTACTACGCGGAGAAATGCCATATGCG 250
Db 90 LysThrAspLysSerGlyProHisSerHisArgLeuTyrTyrIleuGlyMetProTyrGly 109
QY 251 AGCCGAGAGACTCCCTCTCTACTGTGAGATTCCCAAGAGTCCGGAAGGCTCTG 310
Db 110 SerArgGlnAsnSerIleuLeuTyrSerGluIleProLysValArgLysGlnAlaLeu 129
QY 311 CTGCTCTGTCTCGGAGGAGATGCTGTGATCATTTTCCAGGCCACGCCCACTGGGCTC 370
Db 130 LeuLeuLeuSerTyrLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 149
QY 371 TACTCTCGGAGAGGAGCTGTGAGGAGCGGAAACGCTGGGGGTCTTCGCGATCAC 430
Db 150 TyrSerArgGlnGlnGlnLeuLeuArgGlnArgLysArgLeuGlyValAlaPheGlyIleThr 169
QY 431 TCCCTACGACTTCCACAGCAGAGTGGCTTCTCTTCCAGGCGACCAAGGCTCTTC 490
Db 170 SerTyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
QY 491 CACTGTGCGCAGCGGCGGCAAGACGGCTTCAATGTTCCCTTGAACCGCTTGAATC 550
Db 190 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 209
QY 551 AAGACCGAGTGTCCAGGCGCCCGGATGAGCCCAAAATTCCTGCGCGACCGCTTC 610
Db 210 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
QY 611 TTCTCTTCATCATTAACACAGCACTGTGGTGCGCAACATCGACAGCGAGGAGCGG 670
Db 230 PheSerPheIleAsnAsnSerAspLeuTyrValAlaAsnIleGlnIleThrGlnGlnArg 249
QY 671 CGGCTGACCTTTCGCCACCAAGGTTTATCCAAATGCTCTGGATGACCCCAAGTCTGCGGCT 730
|||||
```

Db 250 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspProLysSerAlaGly 269
 QY 731 GTGGCAACCTGTCATPACAGAGAGTTGACCGCTTCACCTGGGTACTGGTGGGCC 790
 Db 270 ValAlaThrPheValIleGlnGlnLeuPheAspArgPheThrGlyTyrTrpPro 289
 QY 791 ACAGCCTCTGGGAAGTTTCAGAGGGCTCAAGAGCTGCAGAACTCTGATAGAGAGTC 850
 Db 290 ThrIleSerTrpGlnGlySerGlnGlyLeuSerThrIleAsnGlyIleLeuTyrGlnVal 309
 QY 851 GATAGATCCGAGGTGGAGGTATTCAGTCCCTCTCTCGCTGCTAGAAAGAAAGACG 910
 Db 310 AspGlnSerGlnValIleGlnValIleHisValProSerProAlaLeuGlnGlyAsnThr 329
 QY 911 GACATGATGCTGATCCCGAGACAGGACAGAAATCCCAAGATTGCTTGAACCTGGCT 970
 Db 330 AsperTrpArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
 QY 971 GAGTTCCAGACTGACAGCCAGGCGCAAGATCGTCTGACCCAGAGAAAGAGCTGGTCAG 1030
 Db 350 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnGlnLeuValGln 369
 QY 1031 CCCTTCAGCTGCTGTTCCCAAGGTGAGTACATGCCAGGCGCGGTGACCCGGAT 1090
 Db 370 ProPheSerSerLeuPheProLysValGlnTyrIleAlaArgAlaGlyTrpThrArgAsp 389
 QY 1091 GGCAAAATAGCGCTGGCGCATTTCTTGACCGGCGCCAGCATGGCTCCACTGCTCTC 1150
 Db 390 G1yLysTyrAlaTrpAlaMetPheLeuAsnArgProGlnGlnTrpLeuGlnLeuValLeu 409
 QY 1151 CTCCCGCGCGCTGTTTCATCCCGACAGACAGAAATGAGAGAGCGCTACGCTCTGCC 1210
 Db 410 LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGlnGlnArgLeuAlaSerAla 429
 QY 1211 AGAGCTGTCCCGAGAGATGTCACCGCTATGTGTGTACGAGAGAGTACCAGCTGTGG 1270
 Db 430 ArgAlaValProArgAsnValGlnProTyrValValTyrGlnGlnValThrAsnValTrp 449
 QY 1271 ATCAATGTCATGACATCTTATCCCTTCCCGCAATCAGAGGGAGAGAGAGAGCTGTGCC 1330
 Db 450 IleAsnValHisAspIlePheThrProPheProGlnSerGlnGlnGlnAspGlnLeuLys 469
 QY 1331 TTTCCTCGCGCGCAATGATGACAGACCGCTTCTGCATTTGTACAAAGTACACCGCTT 1390
 Db 470 PheLeuArgAlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 489
 QY 1391 TTTAAATCCAGGGCTAGATTTGAGTAGCCCTTACGCGCGCGGGAAGATGAATTAAAG 1450
 Db 490 LeuLysSerGlnGlyTyrAspTrpSerGlnProPheSerProGlnGlnAspGlnPheLys 509
 QY 1451 TGCCCATTAAGGAAGATGCTGTGACAGCGGTGAAATGGAGGTTTGGCGAGGAC 1510
 Db 510 CysProIleLysGlnGlnIleAlaLeuThrSerGlnGlnGlnIleValLeuAlaTrpHis 529
 QY 1511 GGCTCCAGAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGACCAAGAGAC 1570
 Db 530 GlySerLys-----GlyThrLysAsp 536
 QY 1571 ACGCGCGTGGAGACACACCTTACGTGTCACATAGAGGGCGCGGAGAGATCTGACGC 1630
 Db 537 ThrProLeuGlnHisHisLeuTyrValValSerTyrGlnAlaIleAlaGlnIleValArg 556
 QY 1631 CTCACGACGCGCGGTCTCCCATAGCTGCTCCATGACGACCGAATCTGACATGTTGCTGC 1690
 Db 557 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspPheVal 576
 QY 1691 AGCCACTACAGACAGGTGAGACAGCGCGCTGCGTGCACAGTCTACAGCTGACGCGCCC 1750
 Db 577 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerLysPro 596
 QY 1751 GACGACGACCGCTGACACAGAGCGCGCTTCTGGGTAGCATGATGAGAGCGACCGAC 1810
 Db 597 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGlnAlaIleAsn 616

QY 1811 TGCCCCCGAGATTATGTTCTCTCAGAGATCTTCATTTCCACAGCGGCTGGATGGCG 1870
 Db 617 CysProProAspTyrValProProGlnIlePheHisPheHisThrAlaGSerAspAlaArg 636
 QY 1871 CTTCAGCGCATGATCTACAAAGCCCCACGCTCTGACCGCGGGAAGAAACACCCACGCTC 1930
 Db 637 LeuTyrGlnMetIleTyrLysProHisAlaLeuGlnProGlnGlnLysHisProThrVal 656
 QY 1931 CTCTTTGATATGAGGCCCCCAGGTGCAGCTGGTGAATTAACCTCTTAAGGATCAAG 1990
 Db 657 LeuPheValTyrGlnGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 676
 QY 1991 TACTTGGCGCTCAACACACCTGCGCTCCCTGGGCTACGCGCGGTGTGATGAGCGGACG 2050
 Db 677 TyrLeuAlaGlyLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValValIleAspValArg 696
 QY 2051 GGCTCTCTGACGCGAGGCTTCTCGGTTCCAGAGGGCGCTGAAAACCAATGGCGCAGGTG 2110
 Db 697 GlySerCysGlnArgGlyLeuArgPheGlnGlnGlnAlaLeuLysAsnGlnMetGlyGlnVal 716
 QY 2111 GAGATCGAGACACAGGTGAGAGCGCTGCAATTCTGTCGCGAAGATAGCTTCATCGAC 2170
 Db 717 GluIleGlnAspGlnValGlnGlyLeuGlnPheValAlaGlnLysTyrGlyPheIleAsp 736
 QY 2171 CTGAGCGAGTTGCCATCCATGGCTGTCGTCATACGGGGGCTCTCTGCTCATGGGGCTA 2230
 Db 737 LeuSerArgValAlaIleHisGlyTrpSerTyrGlnGlyPheLeuSerLeuMetGlyLeu 756
 QY 2231 ATCCACAAGCCCCAGGTGTTCAAGGTGCCATCCGCGGGTGGCCCGCTGACCTGCGATG 2290
 Db 757 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMet 776
 QY 2291 GCCTACGACACAGGTGATACGTAGCGCTACATGACAGTCCCTGAGAACACACGACGCGC 2350
 Db 777 AlaTyrAspTrpGlnGlyTyrThrGlnArgTyrMetAspValProGlnAsnAsnGlnHisGly 796
 QY 2351 TATGAGGGGGGTTCGCTGGCGCTGACGAGGAGAGCTGACCAAGACCCCAACGCGCTG 2410
 Db 797 TyrGlnAlaGlySerValAlaLeuHisValGlnLysLeuProAsnGlnProAsnArgLeu 816
 QY 2411 CTTATCTCTCAGCGCTCTCTGACGAAACAGCGCACTTTTCCACACAACCTCTCTGCTC 2470
 Db 817 LeuIleLeuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuVal 836
 QY 2471 TCCCACTGATCCGACAGGGAACCTTACACAGCTCCAGATCTACCCACAGAGACAC 2530
 Db 837 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnArgHis 856
 QY 2531 AGTATGCTGCGCGCGAGTCCGGGCGAGACATGAAAGTCAGCTGCTGACACTTCTACAG 2590
 Db 857 SerIleArgCysProGlnSerIleGlnHisTyrGlnValThrLeuLeuHisPheLeuGln 876
 QY 2591 GAATACCTC 2599
 Db 877 GluTyrLeu 879
 RESULT 7
 AAE24171
 ID AAE24171 standard; Protein: 830 AA.
 AC AAE24171;
 DE 23-SEP-2002 (first entry)
 DE Human dipeptidyl peptidase 4 (DPP4)-like 2 protein.
 DE Human: dipeptidyl peptidase: DPP: neoplasia: type II diabetes: cirrhosis:
 KW autoimmunity: human immuno deficiency virus: HIV infection: cytostatic:
 KW graft rejection: antidiabetic: antinflammatory: immunosuppressive:
 KW antiviral: enzyme: DPP-4 like 2 protein.
 OS Homo sapiens.

Key	Location/Qualifiers
Key	Misc-difference 235
FT	/note- "Encoded by GAG"
XX	WO200234900-A1.
XX	02-MAY-2002.
XX	29-OCT-2001; 2001WO-AU01388.
XX	27-OCT-2000; 2000AU-0001078.
XX	(UNSY) UNITV SYDNEY.
XX	Abbott CA, Gorrell MD;
XX	WPI: 2002-454646/48.
XX	N-PSDB; AAD38957.
XX	New dipeptidyl peptidase (DPP) peptides, useful for screening
XX	inhibitors of DPP catalytic activity, which may be employed to treat
XX	e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX	rejection and HIV infection -
XX	Disclosure: Page 82-86; 91pp; English.
XX	The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX	polynucleotides encoding such proteins. The DPP peptides are useful for
XX	screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX	for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX	rejection and HIV (human immunodeficiency virus) infection. The present
XX	sequence is human DPP4-like 2 protein.
SQ	Sequence 830 AA:
Alignment Scores:	
Pred. No.:	0
Score:	4458.00
Percent Similarity:	99.88%
Best Local Similarity:	99.76%
Query Match:	92.97%
DB:	23 Gaps: 0
US-09-976-674-4 (1-2617) x AAE24171 (1-830)	
QY	110 CTTCCGGAGCATCATCCAGCGCAGCCGCAAGTACTCGGCGCTATTGCAACAGCGGCC 169
Db	1 LeuAlaGSerLeIleHISglSerAlaGlySTySerIleuIleValaSLysAlaPro 20
QY	170 CAGCACTTCAGATTGTGCGAAGACGGAGATGACTGTGGGCCCACTCCAGCGCTTAC 229
Db	21 HisAspRheGlnPheValaGlnLysTrpAspGluSerGlyProHisSerHisArgLeuTyr 40
QY	230 TACCTGGGAATGCCATATGGCGACGCCAGAGAACTCCCTCTACTGTGAGATTCCCAAG 289
Db	41 TyrLeuGlyMetSerProTyrGlySerArgLysAsnSerLeuLeuTyrSerGlnIleProLys 60
QY	290 AAGGTCCGGAAAGAGCGCTGTGCTGCCTCTCTCTCGAAGAGATGCTGGATCATTTCCAG 349
Db	61 LysValaArgLysGlnAlaLeuLeuLeuLeuSerTTrpLysGlnMetLeuAspHisPheGln 80
QY	350 GCCACGCCCCACCATGAGGGGTCTACTCTCGGAGAGAGAGACTGCTGAGGAGCGGAACGC 409
Db	81 AlaTrpTrpHisHisIleValaIlyrSerArgGlnGlnIleuLeuAlaGlnAlaGlyLysArg 100
QY	410 CTGGGGGCTTCTGGGCACTCACTCTCAAGATTCCACAGCAGAGATGGCTTCTCTCTTC 469
Db	101 LeuGlyValaPheGlyLLehrTrSerTyrAspRheHisSerGluSerGlyLeuPheLeuPhe 120
QY	470 CAGGCGAGACAGCCTTTCCACTCTCGCGAGCGGCGGCAAGACGAGCTTCACTGCTGCC 529
Db	121 GlnAlaSerAsnSerLeuPheHisCysAlaGlySGlyValLysAsnGlyPheMetValSer 140

[illegible]

QY 1610 GCGGCGGCGAGATGTACGGCTTCACAGCGCCGGCTTCCCATAGCTGCATGAGC 1669
 |||
 Db 501 AlAlaIaGIuIleValArgLeuThrThrProGIyPheSerHisSerCysSerMetSer 520
 QY 1670 CAGAACTTCAGCATGTGTGACGCACTACAGACGGTGAGACCGCCGCTGCGTGAC 1729
 |||
 Db 521 GlnAsnPheAspMetPheValSerHisTyrSerSerValSerThrProCysValHis 540
 QY 1730 GTCAACAAGCTGAGCGCCCGGAGAGACGCCCTGCACAGACGAGCCCGCTTGCGGCT 1789
 |||
 Db 541 ValTyrTrpLeuSerGlyProAspAspProLeuHisLysGlnProArgPheThrPala 560
 QY 1790 AGCATGATGAGAGCGACGACGCTGCCCGCGATTATGTCCTCCAGAGATGTCATGTC 1849
 |||
 Db 561 SerMetMetGluAlaAlaSerCysProProAspTyrValProProGluIlePheHisPhe 580
 QY 1850 CACACGGCGCTCGATGTGCGGCTTCAGGCAATGATCAAGACCCCGCCTTGACGCCA 1909
 |||
 Db 581 HisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnPro 600
 QY 1910 GGGAAAGAGACACCCCGCTCTCTTGTATATGAGAGCGCCCGAGGTGACAGTGAAT 1969
 |||
 Db 601 GlyLysLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAsn 620
 QY 1970 AACTCTTCAAGGCAATCACTAGTGTGCGGCTCAACACACTGCGCTCCCTGCGGTACGCC 2029
 |||
 Db 621 AsnSerPheLysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAla 640
 QY 2030 GTGGTTGTGATGACGCGGCGGCTCTGTACAGGAGGCTTCGGTTGGAAGGGCGCTG 2089
 |||
 Db 641 ValValValIleAspArgLysArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeu 660
 QY 2090 AAAAACCAGATGGCGCGGTGAGATGACAGACAGCGGTGAGGGCTCTCACTGTGCGCC 2149
 |||
 Db 661 LysAsnGlnMetGlyGlnValGlnIleGlnAspGlnValGlnGlyLeuGlnPheValAla 680
 QY 2150 GAGAAGTATGAGCTTCATGACGCTGAGCGGATGCGCATGCGTGTGCTGACGCGGCGC 2209
 |||
 Db 681 GluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTyrPheTyrGlyGly 700
 QY 2210 TTCTCTGCTCATGAGGGCTAATCCACAAAGCCCGAGTGTTCAGAGTGGCATCGCGGCT 2269
 |||
 Db 701 PheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaGly 720
 QY 2270 GCGCGGCTCACCGCTGTGATGAGCTACAGACAGGGGTACATGAGCGCTACATGAGACTG 2329
 |||
 Db 721 AlaProValIleThrValIleTrpMetAlaIleTyrAspThrGlyTyrTrpGluArgTyrMetAspVal 740
 QY 2330 CCTGAGAACACACAGACGAGCTATGAGCGGCTTCGCGGCTGACGCTGAGAGAGCTG 2389
 |||
 Db 741 ProGlnAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeu 760
 QY 2390 CCCAATGAGCCCAACCGCTTCCTTATCCCTCCAGCGGCTTCCTGAGGAAGAGTGCACATT 2449
 |||
 Db 761 ProAsnGlnProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhe 780
 QY 2450 TTCACACAAACTTCCTGCTGTCCCACTGATCCGAGACAGGAGAACTTACAGCTCAG 2509
 |||
 Db 781 PheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGln 800
 QY 2510 ATCAACCCCAAGAGAGACAGATATTCGTGCGCCCGAGTGGGGGAGACACTATGAATGC 2569
 |||
 Db 801 IleLysTyrProAsnGlnAlaGlyHisSerIleArgCysProGlnSerGlyGlnHisTyrGlnVal 820
 QY 2570 ACGTGTGCTGCACTTGTACAGGAATACCTG 2599
 |||
 Db 821 ThrLeuLeuHisPheLeuGlnGlyTyrLeu 830
 RESULT 8
 AAE24169
 ID AAE24169 standard; Protein; 869 AA.
 XX
 AC AAE24169;

XX 23-SEP-2002 (first entry)
 DE Alternative version of murine dipeptidyl peptidase 9 (DPP9) protein.
 XX
 XX Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
 KW human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
 KW type II diabetes; antidiabetic; antiinflammatory; immunosuppressive;
 KW antiviral; enzyme.
 KM
 OS Mus sp.
 PN WO200234900-A1.
 XX
 PD 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-AU01388.
 XX
 XX 27-OCT-2000; 2000AU-0001078.
 XX
 XX (UNSY) UNIV SYDNEY.
 PA Abbott CA, Gorrell MD;
 PI
 XX
 XX WPI: 2002-454646/48.
 DR N-PSDB: MAD38955.
 XX
 PT New dipeptidyl peptidase (DPP) peptides, useful for screening
 PT inhibitors of DPP catalytic activity, which may be employed to treat
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 PT rejection and HIV infection -
 PS
 PS Claim 1: Page 70-74; 91pp: English.
 XX
 XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
 CC polynucleotides encoding such proteins. The DPP peptides are useful for
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 CC rejection and HIV (human immuno deficiency virus) infection. The present
 CC sequence is an alternative version of murine DPP9 protein.
 CC Note: This sequence is stated to be the same as that shown as
 CC SEQ ID NO: 4 in the sequence listing of the specification. However these
 CC sequences differ.
 CC
 XX
 SO Sequence 869 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 869
 Score: 4279.00 Matches: 792
 Percent Similarity: 95.81% Conservative: 32
 Best Local Similarity: 92.09% Mismatches: 36
 Query Match: 89.24% Indels: 0
 DB: 23 Gaps: 0
 US-09-976-674-4 (1-2617) x AAE24169 (1-869)
 QY 20 ACGGAGACCCCAACGCGCGAGCGGACAGCGCCGACAGATGACCCGCGCGCCG 79
 |||
 Db 10 SerGIyAlaSerProValGlnValAlaIaIaGlyAspMetAspThrAlaIaArg 29
 QY 80 TTCACAGTGCAGAACACATCGTGGGAGCGGCTCCGAGACATCAATCCAGAGCGCGAG 139
 |||
 Db 30 PheCysValGlnLysHisSerTyrAspArgLeuArgSerIleIleHisGlySerArgLys 49
 QY 140 TACTCGGAGCTCATGTGTCACAGAGCGCCGACGATTCAGATTGTGCAGAAAGAGGAT 199
 |||
 Db 50 SerSerGlyLeuIleValAlaSerLysAlaProHisAspPheGlnPheValGlnLysProAsp 69
 QY 200 GAGTGTGGGCGCCACTCCACCGGCTCTACTACTGAGAAATGCAATGAGACCGGAGAG 259
 |||
 Db 70 GluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGlySerArgGlu 89
 QY 260 AACTCCCTCCCTACTACTGTGAGATTGCCAAGAAAGTCCGGAAGAGGCTGTGCTGCTG 319


```
|||||
Db 90 AsnSerLeuLeuYrSerGluLeProLysLysValAlrGlySgluAlaLeuLeuLeu 109
320 TCCGGAAGAGATGGATGATTCATTCAGAGCCAGCCAGCCACCATGGGGGTACTCTCGG 379
|||||
Db 110 SerTrpLysGlnMetLeuAsnHisPheGlnAlaThrProHisSglYValTlySerArg 129
380 GAGGAGAGGTGCTGAGGAGACGGAAAGCCCTGGGGGTCTTCGGCATCACCTCTACGAC 439
|||||
Db 130 GluGlnLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThrSerTyrAsp 149
440 TTCACAGCCAGAGTGGCCTTCTCTTCAGAGCCAGCAACACCTCTTCACATGTCTGC 499
|||||
Db 150 PheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerSeriLeuPheHisCysArg 169
500 GAGCGCGCAAGACGGCTTCATGTGTCGCCCTGAAACCGCGGAAATCAAGACCCAG 559
|||||
Db 170 AspGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIleLysThrGln 189
560 TGCACAGGCCCCGAGATGAGCCCAAAATCTGCCCTGCCAGCCCTGCTCTCTCCCTTC 619
|||||
Db 190 CysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPhe 209
620 ATCAATACAGCAACCTGTGGGTGCCACATCGAGACAGCGGAGCGGCGCTGAC 679
|||||
Db 210 IleAsnSeriAspLeuTrpValAlaAsnIleGluThrGlyGluArgLysArgLeuThr 229
680 TTTCGCCACCAAGGTTTATCCAAATGTCTGGATGACCCCAAGTGGGGGGTGGCCACC 739
|||||
Db 230 PheLysHisGlnGlySerAlaGlyValLeuAspAsnProLysSerAlaGlyValAlaThr 249
740 TTGCTACACAGGAAGATTGACCGGCTTCACTGGTACTGGTGGTGGCCACAGCTCC 799
|||||
Db 250 PheValIleGlnGluGluPheAspArgPheThrGlyCysTrpTrpCysProThrAlaSer 269
800 TGGGAAGTTCAGAGGCGCTCAGACCGCTGCAATCTCTGATGAGAAAGTTCGATGATCC 859
|||||
Db 270 TrpGluLysSerGluGlyLeuLysThrLeuArgIleLeuTrpGluGluValAlaSer 289
860 GAGGTGAGGTCATTCACGTCCTCTCTGCGCTAGAGAAAGAAAGAGAGGATCTGAT 919
|||||
Db 290 GluValGluValIleHisValProSerProAlaLeuGluGluValGlyLysThrAspSerTyr 309
920 CGGTACCCAGGAGGACGACCAAGATCCCAAGATCCCTTGAATCGCTGATTCAG 979
|||||
Db 310 ArgTyrProArgTrpArgLysSerLysAsnProLysIleAlaLeuLysIleAlaGluLeuGln 329
980 ACTGACAGCCAGGCAAGATGCTCTGACCCAGAGAAAGAGCTGGTGCAGCCCTTACG 1039
|||||
Db 330 ThrAspHisGlnGlyLysIleValIleSerSerCysGluLysGluLeuValGlnProPheSer 349
1040 TCGGTGTCCGGAAGGTGAGATACATCGCCAGGCGCGGTGGAGCCGCGGATGGCAATAC 1099
|||||
Db 350 SerLeuPheProLysValIleGluTyrIleAlaArgAlaGlyTrpThrAlaGlyAspGlyLysTyr 369
1100 GCCTGGGCAATGTTCTGAGACGGGCCAGCAGTGGCTCCAGCTCTCTCTCCCGG 1159
|||||
Db 370 AlaTrpAlaMetPheLeuAspArgProGlnGlnArgLeuGlnLeuValIleLeuLeuProPro 389
1160 GCCCTGTTCATCCGAGACACAGAAATGAGAGAGCGGCTAGCCTTGCCAGAGCTGTC 1219
|||||
Db 390 AlaLeuPheIleProAlaValIleGluSerGluAlaGlnArgGlnAlaAlaAlaGlyAlaVal 409
1220 CCCAGGAATGTCCAGCGGTATGTGTGTACAGAGAGGTTCACCAACGTCGTGATATGTT 1279
|||||
Db 410 ProLysAsnValGlnProPheValIleTyrGluGluValThrAsnValITrpIleAsnVal 429
1280 CATGACATCTTATATCCCTTCCCAATCAGAGGAGAGAGAGAGCTGTGCTTCTCCGC 1339
|||||
Db 430 HisAspTrpIlePheHisProPheProGlnAlaGluGlyGlnIleAspPheCysPheLeuArg 449
1340 GCCAATGAATGCAAGACCGGCTTCTGCCATTTGTACAAAGTCAACGCGCTTTTAAATCC 1399
|||||
```


OY	2480	ATCCGACAGGAACCTTACAGGTCCAGATCACCACGAAAGACACAGTAATTCGC	2539
Dd	830	Ileatgaaglaaglyllyrprotgclnleuglnltetrrproasnlglatrhiserlltlayr	849
OY	2540	TGCCCGAGTGCGGGAGCACACTATGACAGTTCGCTGCACCTTTCTACAGGAATACCTC	2599
Dd	850	CysatggcglusercglylwhistrglyvalThrleuleuhisPheleucInclnhisleu	869
Result 9			
ID	ABG61605	standard; Protein; 832 AA.	
XX	ABG61605:		
AC			
XX			
DE	12- AUG-2002	(first entry)	
XX			
DE	Human DPRP-2 splice variant #4.		
XX			
KW	Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;		
KM	DPRV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;		
KW	diabetes; infertility; obesity; anorexia; osteoporosis; cancer; stroke;		
KM	heart failure; hypertension; urinary retention; osteoporosis; cancer;		
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;		
KX	dyskinesia; reproductive disorder; inflammatory disorder;		
XX	metabolic disorder.		
OS	Homo sapiens.		
PN	WO200231134-A2.		
PD	18- APR-2002.		
PE	12-OCT-2001; 2001WO-US31874.		
PR	12-OCT-2000; 2000US-240117P.		
PA	(FERR) FERRING BV.		
PI	Qi S, Akinsanya KO, Riviere PJ, Junien J;		
DR	WPI: 2002-444178/47.		
N-P	PSDB; ABRK3336.		
PT	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding		
PT	the proteins, useful for treating e.g. fungal, bacterial, protozoan and		
PT	viral infections, cancers, allergies, neurological disorders, or pain		
PT	.		
PS	Disclosure: Page 85-87; 113pp; English.		
XX			
CC	The present invention relates to the isolation of novel human serine		
CC	proteases referred to as dipeptidyl peptidase IV (DPRIV)-related		
CC	proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)		
CC	and nucleic acids encoding them are useful for treating infections		
CC	such as fungal, bacterial, protozoan and viral infections, particularly		
CC	infections caused by human immunodeficiency virus (HIV-1 or HIV-2),		
CC	pain, diabetes, precocious puberty, infertility, obesity, anorexia,		
CC	bulimia, Parkinson's disease, acute heart failure, hypotension,		
CC	hypertension, urinary retention, osteoporosis, angina pectoris,		
CC	stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,		
CC	psychotic and neurological disorders (e.g. anxiety, dementia, or		
CC	schizophrenia), and dyskinesias. These may also be used in discovering		
CC	therapeutic agents for the treatment of reproductive, inflammatory and		
CC	metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.		
SO	Sequence 832 AA;		
Alignment Scores:			
Pred. No.:	0	Length:	832
Score:	4259.50	Matches:	802
Percent Similarity:	94.13%	Conservative:	0
Best Local Similarity:	94.13%	Mismatches:	1

Query Match:	88.83%	IndeIs:	49
DB:	23	Gaps:	1
US-09-976-674-4 (1-2617) x ABG61605 (1-832)			
OY	11	ATGGCCACACCGGAGCCCCAACGGCCGACCGAGCGAGCGACGCCGCACAGATGACCCG	70
Db	30	MetalathrthrnglythrprothrAlaAspArgGLyAspAlaAlaIalathrsprspro	49
OY	71	GCCGCCGCTTCCAGGTGCAAAAGACCTGTGGAGGGGCTCCGGAGTCATCCAGGC	130
Db	50	AlaAlaArpPheglInValglnLysHisSerTrpAspGLyLeuArgSerIleIleHISGLy	69
OY	131	AGCCGCAAGTCTGGGGCCTCATTTGTACAAGAAGCCCCCAGACTCCAGTTTGTCAG	190
Db	70	SerArgLysTrpSerGLyLeuIleValAsnLysAlaProHisAspPheglInPheValgln	89
OY	191	AAGAGGATGAGTGTGGGGCCCATCCAGCCGCTCTACTACCTGGGAATGGCATATGAC	250
Db	90	LysThrAspArgLysSerGLyProHisSerHisArgLeuTyrrTLyeuLyIweTrotyrGLy	109
OY	251	AGCCGAGAACTCCCTCTACTACTGTGATGCCAATAAGGTCCGGAAGAAGCTCTG	310
Db	110	SerArgLysAsnSerLeuLeuTyrrSerGluILeProLyLysValAlaArgSLlunIaleu	129
OY	311	CTGCTCCTGTCTTGGAAAGCAATGCTGATCATTTCCAGGCCACGCCCATGAGGCTC	370
Db	130	LeuLeuLeuSerTrpLysGLImetLeuAsnHisPheglInAlatrProHisLSGLyAl	149
OY	371	TACTCTGGGAGGAGAGCTGTGAGGAGCGGAAAAGCCTGGGGGTCTTCCGCATCAC	430
Db	150	TyrSerArgGLyugLnugLnLeuLeuArgGLuArgLysArgLengLYAlaPheLIylethr	169
OY	431	TCTTACGACTTCCACACGAGAGTGGGCTTCTCTCTCCAGGCCAGCAAGCCTCTTC	490
Db	170	SerTrpAspPheHisSerGLyLeuPheLeuPheglInAlaserhsSerleuPhe	189
OY	491	CACGTGCGCAGCGCGGCAAGAACGGCTCATGTGTCCTTATGAACCGCTGGAATC	550
Db	190	HIScysArgAspArgLYLysAsnGLyPheMetValSerProMetLysProLeuGLuIle	209
OY	551	AAGACCAGCTCTAGGGCCCCGGATGAGCCCAAATGTGCCCTGCGACCCCTGCTTC	610
Db	210	LysThrGLncysSerGLyProArgMeAspProLysILeLysProAlasProAlaPhe	229
OY	611	TTCCTCTCATCAATTAACAGCAGCTGTGGGTGGCAACATCGACAGCAGCAGAGCGG	670
Db	230	PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnILeGLuthrGLylunGLuArg	249
OY	671	CGGCTGACCTTCTGSCAACAAAGTTTATCCAAATGTCTGTGATGACCCCAAGTCTGGGT	730
Db	250	ArgLeuThrPheCysHisLSGLnGLyLeuSerAsnValIleuAspSprproLysSerAlagly	269
OY	731	GTGGCCACCTTCGTCATACAGAAAGAGATTGGACCGCTCACTAGGGTAGTGGTGGCCC	790
Db	270	ValAlaThrPheValILeGLnugLnugLnPheAspArgPheThrGLyTrpTrpPrysPro	289
OY	791	ACAGCCTCTGGAAAGGTGACAGAGGCTCAACAGCTGGGAATCTGTATAGCAAGTC	850
Db	290	ThraIlaSerTrpGLynGLySerGLynGLyThrLeuArgILeleuTyrrGLunGLuVal	309
OY	851	GATGAGTCCGAGGTGAGGTCAATTCACGTCCCTCTCTGCGTAGAAGAAAGAAAGACG	910
Db	310	AspGLuserrGLuValglnValILeHisValProSerProAlaLeuGLunGLuArgLysThr	329
OY	911	GACTCGTATCGGTACCCGAGACAGGACAGCAAGAATCCCAAGTTTCCCTTGAAC TGCT	970
Db	330	AspSerTrpArgTrpProArgThrGLySerLysAsnProLysILeAlaLeuLysLeuAla	349
OY	971	GAGTCCCGAGTGCAGCAGGCGCAAGATGTCGACCCGAGAGAAAGAGTGGTGGAG	1030
Db	350	GIuPheglInThrAspSerGLnGLyLysILeValSerThrGLnGLnGLySGLuLeuValGLn	369

QY 1031 CCCTTCAGCTCGCTTCCTCCGAGAGGTGAGTACATCCGACGCGGGGTGACCCGGGAT 1090
 |||||||
 Db 370 ProPheSerSerLeuPheProGluValGluTyrIleAlaArgAlaGlyTrpThrArgasp 389
 |||||||
 QY 1091 GGCAGAAATACGCGTGGCATGTCTCTGACCGCGCCAGCATGGCTCCACCTGCTCTC 1150
 |||||||
 Db 390 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 409
 |||||||
 QY 1151 CTCGCCCGGCGCTTCATCCCGACAGACAGATGAGAGGACGAGCGCTACCTGCTGCC 1210
 |||||||
 Db 410 LeuProProAlaLeuPheIleProSerThrGluAsnGluGlnAlaGlyLeuAlaSerAla 429
 |||||||
 QY 1211 AGAGCTCTCCCGAGGAATGTCCAGCCGTATGTGTGACGAGAGGTCCACCACTGTGG 1270
 |||||||
 Db 430 ArgAlaValProArgAsnValGlnProTyrValAlaTyrGluGlnValThrAsnValTrp 449
 |||||||
 QY 1271 ATCAATGTATGACATCTTATCCCTTCCCAATCCAGGGGAGAGAGACGACTCTGC 1330
 |||||||
 Db 450 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGlnAspGluLeuGln 469
 |||||||
 QY 1331 TTTCTCCGCGCAATGATGATGACAGACGCGCTTTCGCAATTTGTACAAAGTACCGCGCT 1390
 |||||||
 Db 470 PheLeuArgAlaAsnGluGlnGlySerThrGlyPheCysHisLeuTyrValThrAlaVal 489
 |||||||
 QY 1391 TTTAAATCCAGGCGTACGATGTGAGTGGAGCCCTTCAGCCCGGGGAGATGAATTTAG 1450
 |||||||
 Db 490 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGlyGluAspGluPheLys 509
 |||||||
 QY 1451 TGGCCCATTTAAGAGAGATTTGCTGTACACGCGGTAAATCGGAGGTTTGGCCAGGCAC 1510
 |||||||
 Db 510 CysProIleLysGluGlnIleAlaLeuThrSerGlyLeuTrpGluValLeuAlaArgHis 529
 |||||||
 QY 1511 GCGTCCAAAGTCTGGGTCAATGAGAGAGACCAAGCTGGTGTACTTCCAGGCGACCAAGAC 1570
 |||||||
 Db 530 GlySerLysIleTrpValAsnGluGlnThrLysLeuValTyrPheGlnGlyThrLysAsp 549
 |||||||
 QY 1571 AGCGCGGTGAGACACACCTCTACGTGTGACGTATAGAGCGCGCGAGANTGTACGC 1630
 |||||||
 Db 550 ThrProLeuGlnHisHisLeuTyrValValSerTyrGluAlaGlnGlyLeuValArg 569
 |||||||
 QY 1631 CTACACGACGCGCGCTTCCCATAGTGTGCTGCATGAGCCAGAACTTCGACATTTGCTC 1690
 |||||||
 Db 570 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 589
 |||||||
 QY 1691 AGGCATPACAGACAGCTGAGACAGCGCGCGCTGTGACAGCTTCAACAGTGAAGCGCGCCC 1750
 |||||||
 Db 590 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 609
 |||||||
 QY 1751 GAGGAGACGCGCGCTGCAAGAGACCGCGCTTGGGTGACATGATGAGAGCGCGAC 1810
 |||||||
 Db 610 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetGluAlaAlaSer 629
 |||||||
 QY 1811 TGGCCCCGGATTATGTTCTCCAGAGATCTTCATTTCCACACGCGCTGGAGTGTGCGG 1870
 |||||||
 Db 630 CysProProAspTyrValProProGlnIlePheHisPheHisThrArgSerAspValArg 649
 |||||||
 QY 1871 CTCTAGGCGATGTATPACAAAGCCCAAGCCTTGACAGCCAGAGAAAGACACCCACGCTC 1930
 |||||||
 Db 650 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal 669
 |||||||
 QY 1931 CTCTTTGTATATGAGAGCGCGCGAGGTGACGTGGTGAATATCTCTTCAAGGCATCAAG 1990
 |||||||
 Db 670 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 689
 |||||||
 QY 1991 TACTTGCGGTCAACACAGCGGCTCCGTGGGTACGCGGTGGTGTGATGAGCGAGG 2050
 |||||||
 Db 690 TyrLeuThrGluLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArg 709
 |||||||
 QY 2051 GGCCTCTGTACAGAGGCGCTTCGTTTCGAAGGCGCGCTTCAAAAACCAATAGCGCAGGTG 2110
 |||||||
 Db 710 GlySerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnVal 729
 |||||||
 QY 2111 GAGATGAGAGACCAAGGTGAGGCGCTGCAGTTCGTGGCCGAGAAAGTATGGCTTCATCGAC 2170
 |||||||

Db 730 GluIleGluAspGlnValGluGlnIlePheValAlaGluLysTyrGlyPheIleAsp 749
 |||||||
 QY 2171 CTGAGCGGAGTGTGGCATTCATAGCGTGTCTACAGGGGCGTCTCTGCTCATGGGCTA 2230
 |||||||
 Db 750 LeuSerArgValAlaIleHisGlyTyrPheSerTyrGlyGlyPheLeuSerLeuMetGlyLeu 769
 |||||||
 QY 2231 ATCCACAAGCGCGAGGTTCACAGGTGGCCATCGCGGGTGGCCCGGTCACCGTCTGGATG 2290
 |||||||
 Db 770 IleHisLysProGlnValPheLys--Ala----- 778
 |||||||
 QY 2291 GCCTAGACACAGGGGTACTGAGCGCTACATGAGAGCTCCCTGAGAACACACAGCAGCGC 2350
 |||||||
 Db 778 ----- 778
 |||||||
 QY 2351 TATGAGCGGGTTCCGTGCGGCTCTCAGGTGAGAGAGCTGCCCAATGAGCCCAACCGCTTG 2410
 |||||||
 Db 779 -----GlnProLeuA 782
 |||||||
 QY 2411 CTATTCCTCCAGCGCTTCCTGAGACGAAACGTGACACTTTTCCACAGAACTTCTGCTC 2470
 |||||||
 Db 782 IatYrProProArgLeuProGlyArgLysArgAlaLeuPheProHisLysLeuProArgL 802
 |||||||
 QY 2471 TCCCACTGATCCGAGGAGGAAACCTTACCAGCTCCAGATCTACCCCAAGAGAGACAC 2530
 |||||||
 Db 802 euProThrAspProSerArgGluThrLeuProAlaProAspLeuProGlnArgGluThrG 822
 |||||||
 QY 2531 AGTATTCGCTGCGCGCGAGTGGCGGAGCAGCTA 2562
 |||||||
 Db 822 IatYrSerLeuProArgValGlyArgAlaLeu 832
 |||||||
 RESULT 10
 ABG61606
 ID ABG61606 standard; protein; 832 AA.
 XX
 AC ABG61606;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-2 splice variant #5.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinnesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 XX
 OS Homo sapiens.
 OS
 PN WO200231134-A2.
 PN
 PD 18-APR-2002.
 PD
 XX 12-OCT-2001; 2001WO-US31874.
 PF
 XX 12-OCT-2000; 2000US-240117P.
 PR
 XX (FERR) FERRING BV.
 PA
 XX OI S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 PI WPI: 2002-444178/47.
 DR N-PSDB; ABR83337.
 DR
 XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 XX
 PS Disclosure: Page 88-90; 113pp; English.

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPIV)-related
 CC and nucleic acids encoding them are useful for treating proteins (DPRP)
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 832 AA;

Alignment Scores:

Pred. No.: 0 Length: 832
 Score: 4259.50 Matches: 802
 Percent Similarity: 94.13% Conservative: 0
 Best Local Similarity: 94.13% Mismatches: 1
 Query Match: 88.83% Indels: 49
 DB: 23 Gaps: 1

US-09-976-674-4 (1-2617) x ABG61606 (1-832)

OY 11 ATGGCCACACCGGAGCCCAAGCGCGAGCGAGCGAGCCGACAGATGACCCG 70
 Db 30 MetalafrhrhnglythrProthrAlaaspraglyAspAlaAlaAlaAlaThrAspAspPro 49
 OY 71 GCGCGCCGCTTCAGGTCAGAGCAGCTGTCGGGCGGCGGCTCCGAGATCATCCAGGCG 130
 Db 50 AlaAlaAlaArgheglnValglnLysHisSerThrAspArglyLeuArgSerIleHisgLy 69
 OY 131 AGCGGCAAGTACTCGGCGCTCATTTGTCAACAAGCGCGCCCAAGATTCAGATTGTGAG 190
 Db 70 SerArgLysTyrSerGlyLeuIleValAlaLysAlaProHisAspPheGlnPheValgln 89
 OY 191 AAGACGAGTACGTGGGCGCCCACTCCACGCGCTTACTACCTTGGAATGCCATATGCG 250
 Db 90 LysThrAspArgLysSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 109
 OY 251 AGCGGAGAGTCCCTCTACTCTGAGATTCACCAAGAGGTCGCGAAGAGGCTCG 310
 Db 110 SerArgLysAsnSerLeuLeuTyrSerGlyIleProLysValArgLysGlnAlaLeu 129
 OY 311 CTGCTCTGTCTGTGAGACAGATGCTGATCATTTCCAGGCGACGCCACCATGGGCTC 370
 Db 130 LeuLeuLeuSerThrPryGlnMetLeuAspHisPheGlnAlaThrProHisHisgLyAl 149
 OY 371 TACTCTCGGAGAGGAGACTGTGAGGAGCGGAACGCGTGGGCTTCCGCGATCAC 430
 Db 150 TyrSerArgLysGlnGlnLeuLeuArgLysArgLeuGlyValPheGlyIleThr 169
 OY 431 TCCATACGACTTCCACAGGAGAGTGGCTTCCTCTCCAGGCGAGCAACGCGCTTTC 490
 Db 170 SerTyrAspPheHisSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
 OY 491 CACTTCGCGAGCGGCGCAAGACGGCTTTCATGTGTGCTCCCTATGAACCGCTGGAATC 550
 Db 190 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuIle 209
 OY 551 AAGACCAAGTGTCTAGGCGCCCGATGAGACCCCAAAATTCGCTGCGACCGCTTCC 610
 Db 210 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
 OY 611 TTCCTCTCATCAATAACAGCGACTGTGGGCGCAACATGAGACAGCGGAGAGAGCG 670
 Db 230 PheSerPheIleAsnAsnSerAspLeuThrValAlaAsnIleGlnThrGlyGlnIleArg 249
 OY 671 CGGCTGACCTTCTGCGACCAAGGTTATCAATGTCTGATGACCCCAAGTCTCGGCT 730

Db 250 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 269
 OY 731 GTGGCCACCTTCGTCATACAGAGAGTTCGACCGCTTCACTGGGTACTGGTGGTGGCC 790
 Db 270 ValAlaThrPheValIleGlnGlnLeuIlePheAspArgPheThrGlyTyrThrProLysPro 289
 OY 791 ACAGGCTCTGGGAGAGTTCAGAGGCGCTCAAGACGCTGCGCAATCCTATGAGAGATC 850
 Db 290 ThrAlaSerThrPrglnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnVal 309
 OY 851 GATGAGTCCGAGGTGAGGTATTCACGTCCCTCTCGCGCTAGAGAAAGAGAGAGC 910
 Db 310 AspGlnSerGlnValGlnValIleHisValProSerProAlaLeuGlnGlnArgLysThr 329
 OY 911 CACTGTGATCGGTATCCCGACAGAGCGAGCAAGAATCCCAAGATTGCTTGAACGTGCT 970
 Db 330 AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
 OY 971 GAGTTCAGACTGACAGCGAGCGCAAGATCGTCTGACCCAGAGAGAGAGCTGGCGAG 1030
 Db 350 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnLysGlnLeuValgln 369
 OY 1031 CCTTTCAGCTCGCTTCCCGAAGGTGAGTATCATCGCCAGGCGCGGTGACCCGGGAT 1090
 Db 370 ProPheSerSerLeuPheProLysValGlnTyrIleAlaArgAlaGlyTyrPheThrArgAsp 389
 OY 1091 GGCAAAATACGCTGGGCGCATGTCTGTGACCGCGCCCAAGAGTGGCTCCAGTGTCTTC 1150
 Db 390 GlyLysTyrAlaTyrPheAlaMetPheLeuAspArgProGlnGlnThrLeuGlnLeuValLeu 409
 OY 1151 CTCCCGCGCGCGCTGTGATCCCGAGCAGAGATGAGAGAGCAGCGCTGAGCTCGCC 1210
 Db 410 LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGlnArgLeuAlaSerAla 429
 OY 1211 AGAGCTGTCCCGAGAGATGTCCAGCGGATGTGTGTGTCAGAGAGTCAACAGCTGTGG 1270
 Db 430 ArgAlaValAlaProArgAsnValGlnProTyrValIleTyrGlnGlnValThrAsnValThr 449
 OY 1271 ATCATATGTTTCATGACATGTTCTTATCCCTTCCCGCAATCAGAGGAGAGAGAGAGCTTGC 1330
 Db 450 IleAsnValHisAspIlePheTyrProPheProGlnSerGlnGlyLysAspIleLeuLys 469
 OY 1331 TTCTCCGCGGCAATGATGCAAGACCGGCTTCCATTTGTACAAAGTCAACGCGCTT 1390
 Db 470 PheLeuArgAlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 489
 OY 1391 TTAATAATCCAGGCTAGCATTTGAGTGAAGCCCTTCAAGCCCGGAGAGATGAATTAA 1450
 Db 490 LeuLysSerGlnGlyTyrAspThrPserGlnProPheSerProGlyLysAspIlePheLys 509
 OY 1451 TGCCCATTTAAGAGAGATTTGCTGTGACACGCGGTGAATGGAGAGTTTGGCGAGGCAC 1510
 Db 510 CysProIleLysGlnGlnIleAlaLeuThrSerGlyGlnTyrPheLysValLeuAlaArgHis 529
 OY 1511 GGCCTCAAGATGGGTCAATGAGAGACCAAGCGGTGTACTTCCAGGCGCCCAAGAGAC 1570
 Db 530 GlySerLysIleThrPyrAlaAsnGlnIleLysLysLeuValTyrPheGlnGlyThrLysAsp 549
 OY 1571 ACGCGCTGTGAGACACACCTTACTGTGTGATGAGGCGCGCGAGATGCTAGAC 1630
 Db 550 ThrProLeuGlnHisHisLeuTyrValIleSerTyrGlnAlaIleGlyIleValIleArg 569
 OY 1631 CTCACACAGCGCGCGCTTCTCCATACGTGCTCATGAGGCGAGACCTTGACATGTTCTGTC 1690
 Db 570 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 589
 OY 1691 AGCCACTACAGAGGCTGTGAGCGCGCGCTCGTGCAGAGTGTACAGGTGAGCGCGCC 1750
 Db 590 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerIlePro 609
 OY 1751 GACGACAGACCCCTGACACAGAGCGCCGCTTCTGGGCTAGCATGATGAGGACAGCGAC 1810
 Db 610 AspAspAspProLeuHisLysGlnProArgPheThrAlaSerMetMetGlnAlaIleSer 629

```
QY 1811 TGCCCCGGATATGTTCTCCAGATCTTCATTTCCACACGCGCTCGATGTGGG 1870
    |||||||
Db 630 CysProProAspArgValProProGluIlePheHisPheHisSThrArgSerAspValArg 649
QY 1871 CTCACGGCATGATCTACACAGCCCGCTTGACAGCCAGGAAGACACCCAGCTC 1930
    |||||||
Db 650 LeuArgIleMetIleTyrLysProHisAlaLeuInProGlyLysHisProThrVal 669
QY 1931 CTCCTTGATATGGAGGCGCCCGAGTGCAGCTGGGATATCTCTTCAAGGATCAG 1990
    |||||||
Db 670 LeuPheValIleTyrGlyGlyProGluValGluLeuValAsnAsnSerPheLysGlyIleLys 689
QY 1991 TACTGGGCTCAACACACTGGCTCCCTGGGCTACCGCGCTGATTGATTCAGCGCAG 2050
    |||||||
Db 690 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValIleAspGlyArg 709
QY 2051 GGCCTCTGTCAGCGAGGCTTCGGTTCGAAGGGCCCGGAAAAACCAATGGCCAGTG 2110
    |||||||
Db 710 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnVal 729
QY 2111 GAGATTCAGACAGCATGTGGAGGCGCTGCAGTTCGTGGCCGAGAGTATGGCTTCAGAC 2170
    |||||||
Db 730 GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 749
QY 2171 CTGAGCCGAGTTCGCTATCCATGCGTGTCTACAGGGGGCTTCCTGCTCATGGGCTA 2230
    |||||||
Db 750 LeuSerArgValAlaIleHisGlyTyrPserTyrGlyGlyPheLeuSerLeuMetGlyLeu 769
QY 2231 ATCCACAGACCCCGAGTTCCTTAAGGTGGCCATCCGCGGGTGCCTGCTGATG 2290
    |||||||
Db 770 IleHisLysProGlnValPheLys--Ala----- 778
QY 2291 GCTACAGACACAGGCTACACTGACGCGCTACATGACGTCCTGAGAACACACAGCAGCGC 2350
    |||||||
Db 778 ----- 778
QY 2351 TATGAGCGGGTTCCTGGCCCTGCACGTGGAGAGCTGCCCAATGAGCCCAACCGCTTG 2410
    |||||||
Db 779 -----GlnProLeuA 782
QY 2411 CTATTCCTCCACGGCTCTCTGGAGAAACGTGCATTTTCCACACAAACTCTCTGTC 2470
    |||||||
Db 782 IatYrProProAlaArgLeuProGlyArgLysArgAlaLeuPheProHisLysLeuProArgL 802
QY 2471 TCCCAACTGATCCGAGCAGGAGAAACCTTACAGCTCCAGATCTACCCCAACGAGAGACAC 2530
    |||||||
Db 802 eupProThrAspProSerArgGluThrLeuProAlaProAspLeuProGlnArgLThrArg 822
QY 2531 AGTATTCGCTGCCCCGAGTGGGGGAGACACTA 2562
    |||||||
Db 822 IntYrSerLeuProArgValGlyArgAlaLeu 832
RESULT 11
ABG61609
ID ABG61609 standard; Protein: 819 AA.
XX
AC ABG61609;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPPP-2 splice variant #8.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPP;
KW DPPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinestia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
```

```
PN W0200231134-A2.
XX
PD 18-APR-2002.
XX
PE 12-OCT-2001; 2001MO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI: 2002-444178/47.
DR N-PSDB; ABK83340.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the protein, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
PS Disclosure: Page 98-100; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPPIV)-related
CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPPP proteins.
XX
SQ Sequence 819 AA;
XX
Alignment Scores:
Pred. No.: 0 Length: 819
Score: 4172.00 Matches: 789
Percent Similarity: 92.61% Conservative: 0
Best Local Similarity: 92.61% Mismatches: 1
Query Match: 87.01% Indels: 62
DB: 23 Gaps: 2
US-09-976-674-4 (1-2617) x ABG61609 (1-819)
QY 11 ATGGCCACACAGGGGACCCCAACGGCCGAGCGAGCGACGCCGACAGATGACCG 70
    |||||||
Db 30 MetAlaThrThrGlyThrProThrAlaAspArgGlyLysAlaAlaAlaThrAspAspPro 49
QY 71 GCCCGCCGCTTCACGAGTGCAGAAACACTCGTGGGACCGGCTCCGGAGCATCATCCAGGC 130
    |||||||
Db 50 AlaAlaArgPheGlnValGlnLysHisSerTyrAspIleLeuArgSerIleIleHisGly 69
QY 131 AGCCGCAAGTACTCGGCGCTCATGTGTCAACAAGCGCGCCCAAGCACTCCAGTTGTGTCAG 190
    |||||||
Db 70 SerArgLysTyrSerGlyLeuIleValAlaSnLysAlaThrHisAspPheGlnPheValGln 89
QY 251 AGCCGAGAGAACTCCCTCTCTACTGTGAGATTCACAAAGAGTCCGGAAGAGGCTCTG 310
    |||||||
Db 110 SerArgGlnLysSerLeuLeuTyrSerGluIleProLysValAlaGlyGluAlaLeu 129
QY 311 CTGCTCTGTCCTGAGAGCATGCTGGATCATTTCCAGGCGACGCCCAACCATGGGCTC 370
    |||||||
Db 130 LeuLeuLeuSerTyrPlyGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 149
```

QY 371 TACTCTGGGAGAGAGAGCTCTGAGGAGGAGAAACCGCTGGGGCTCTTGGCATCAC 430
 Db 150 TySerarAgglugluGlutLeuLeuArggluArgLysArgLeuGluValPheGlyIleThr 169
 QY 431 TCCATACGACTTCCACAGGAGAGAGTGGCTCTTCTCTTCCAGGCCAGCAACAGCTCTTC 490
 Db 170 SerTyArSPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
 QY 491 CACTGTGCGCGAGCGGAGAGAGCGGCTTACGTGCTCCCTATGAAACCGCTGGAAATC 550
 Db 190 HisCysArAgpAgglYglYlYsAsnGlyPheMetValSerProMetLysProLeuGluIle 209
 QY 551 AAGACCAAGTGTCTAGGCGCCCGGATGGAGCCCAAAATCTGCCCTGCCAGCTTGCCTTC 610
 Db 210 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
 QY 611 TTCTCCCTTCATCATTAACAGGAGGACTGTGGGTGGCCAAATCAGACAGCGGAGAGCGG 670
 Db 230 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg 249
 QY 671 CGGCTGACCTTCTGCCACCAAGGTTTATCCAAATGCTCGATGACCCCAAGCTGGCGGT 730
 Db 250 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValIleuAspAspProLysSerAlaGly 269
 QY 731 GTGGCAACTTCTCATACAGAGAGAGTTCCAGCCGCTTCACTGGGTACTGGTGGCCCC 790
 Db 270 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTrpTrpProCysPro 289
 QY 791 ACAGCCTTCCTGGAGAGGTTGAGAGGGCTCAAGACGCTGGCAATCTGTATGAGAAATC 850
 Db 290 ThrIleSerTrpGlnGlySerGlnGlyLeuYsThrLeuArgIleLeuTyGluGluVal 309
 QY 851 GATAGAGCCGAGGAGGAGGATTCACGTCCCTCTCTCGGCTAGAGAAAGAAAGAGAGCG 910
 Db 310 AspIleuSerGluValIleGluValIleHisValProSerProAlaLeuGluGluArgLysThr 329
 QY 911 GACTCGATGCGTATCCCAAGAGAGAGCAAGAATCCCAAGATTGCTTGAATCGACT 970
 Db 330 AsperTyArgTrpArgTrpProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
 QY 971 GAGTTCCAGAGTGCAGAGCGGAGAGATGCTCTCGACCCAGAGAAAGAGAGAGCTGTGAG 1030
 Db 350 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 369
 QY 1031 CCCCTTCAAGCTCGCTGTTCCCAAGAGTGGAGTACATCGGACAGGGCGGCTGACCGGGAT 1090
 Db 370 ProPheSerSerLeuPheProLysValGluTrpTrpIleAlaArgAlaGlyTrpThrArgAsp 389
 QY 1091 GGCAAATACGCTGGCGCATGTTCTTGAGCCGCGCCAGCAGTGGCTCCAGCTGCTCCTC 1150
 Db 390 GlyLysTyArgAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 409
 QY 1151 CTCCCGCGGCGCTGTTTCATCCCGACACAGAGAATGAGAGCGGCTAGCCTGTGCC 1210
 Db 410 LeuProProAlaLeuPheIleProSerThrLusnGlnGluGlnAlaArgLeuAlaSerAla 429
 QY 1211 AGAGCTGTCCCGAGGAATGTCACGCTATGTGTTAGAGAGAGGTACCAACAGTGG 1270
 Db 430 ArgAlaValProArgAsnValGlnProTyArgValIleTyrgluGluValThrAsnValTrp 449
 QY 1271 ATCAATGTTCATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGAGAGAGCTGTGC 1330
 Db 450 IleAsnValHisAspIlePheTyProPheProGlnSerGlnGluAspGluIleuLys 469
 QY 1331 TTCTCGCGGCGCATGAGATGAGAACCGGCTTTCGCCATTTGTACAAAGTACCGCGGTT 1390
 Db 470 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyLysValThrAlaVal 489
 QY 1391 TTTAAATCCAGGCGCTACGATTTGAGTGGAGCCCTTCAGCCCGGAGAAATGAAATTTAAG 1450
 Db 490 LeuYsSerGlnGlyTyArgSPTrpSerGluProPheSerProGlyGluAspGluPheLys 509
 QY 1451 TGCCCCATTTAAGAGAGATTTGCTGTACACAGCGGTGAATGGAGGTTTTTGGCAGGCAAC 1510

Db 510 CysProIleLysGlnGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArgHis 529
 QY 1511 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTACTTCCAGGCGCAACAGGAC 1570
 Db 530 GlySerLys-----GlyThrLysAsp 536
 QY 1571 ACGCGCGTGGAGCAACACTTACGTGTGACATGAGGGCGCGGAGAAATGGTACGC 1630
 Db 537 ThrProLeuGlnHisLysLeuTyValValSerTyrgluAlaAlaGlyIleValArg 556
 QY 1631 CTCACACAGCGCGCTTCTCCATAGCTGTCTCATGAGCCAGAACTTCCAGATGTGCTC 1690
 Db 557 LeuThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 576
 QY 1691 AGCCACTACACAGCGTGAAGACGCGCGCTTGCAGCTGCAGCTTCAACAGTGAAGCGCCC 1750
 Db 577 SerHisTySerSerValSerThrProProCysValHisValTyLysLeuSerGlyPro 596
 QY 1751 GAGCAGACCGCCCTGACACAGACCGCGCTTTCGGCTTACCATGATGAGAGCGCACGC 1810
 Db 597 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaAlaSer 616
 QY 1811 TGCCCCCGGATTAATGTTCTCCAGAGATCTTCCATTTCCACAGCGCTCGATGTGCGG 1870
 Db 617 CysProProAspTyArgValProProGluIlePheHisPheHisTrpArgSerAspValArg 636
 QY 1871 CTGACGGCATGATCTACAAAGCCCAAGCCTTGCAGCCAGGAGAAAGAACCCACCGCTC 1930
 Db 637 LeuTyrglyMetIleTyLysProHisAlaLeuGlnProGlyLysHisProThrVal 656
 QY 1931 CTCTTGTATATGGAGCGCCCGAGTGCAGCTGGTGAATTAACCTCTTCAAGGATCAAG 1990
 Db 657 LeuPheValTyrglyGlyProGlnValGlnLeuValIleAsnSerPheLysGlyLys 676
 QY 1991 TACTTGGCGCTCAACACACTGGCTCCCTGGGCTTACGCGCGTGTATGAGAGGCGAGG 2050
 Db 677 TyLeuThrArgLeuAsnThrLeuAlaSerLeuGlyTyArgAlaValIleAspGlyArg 696
 QY 2051 GGCCTCTGTACGCGAGGCGCTTCCGTTGGAAGGGCGCTGAAAAACCAATGGCCAGGTG 2110
 Db 697 GlySerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnVal 716
 QY 2111 GAGATCGAGCAGCAGGTGGAGGGCTCGAGTTCGGCGCGAGTATGGCTTCATCGAC 2170
 Db 717 GluIleGluAspGlnValGlnGlyLeuGlnPheValAlaGluLysTyrglyPheIleAsp 736
 QY 2171 CTGAGCGAGTTGGCATATCATGAGCTGTGCTTACAGGGGGCTTCTCTGCTATGGGCGTA 2230
 Db 737 LeuSerArgValAlaIleHisGlyTrpSerTyrglyGlyPheLeuSerLeuMetGlyLeu 756
 QY 2231 ATCCCAAGCCCCAGGCTTTCAGGTGGCCATCGCGGGTGCSCCGCTGTGATG 2290
 Db 757 IleHisLysProGlnValPheLys--Ala----- 765
 QY 2291 GCTTACGACAGGGTACACTGAGCGGTACATGAGCGTCCCTGAGAACACACAGACGGC 2350
 Db 765 ----- 765
 QY 2351 TATGAGCGGGTTCGTTGCGCCCTGACGTGAGAGAGTCCCAATGAGCCCAACCGCTTG 2410
 Db 766 -----GlnProLeuA 769
 QY 2411 CTATATCTCCACGCGCTTCTCGAGAGAAAGCTGACATTTTCCACACAACTTCTGTGC 2470
 Db 769 IatyrProProArgLeuProGluArgLysArgAlaLeuPheProHisLysLeuProArgL 789
 QY 2471 TCCCAAGTATCCAGAGAGGAGAAACCTTACAGGCTCCAGATCTCCCAAGAGAGAGAC 2530
 Db 789 eupTrpHisPProSerArgGluThrLeuProAlaLapProAspLeuProGlnArgGluThrG 809
 QY 2531 AGTATTCGCTGCCCGAGTGGCGGAGCACATA 2562

DB 809 IntYrSerLeuProAlrGValGIyArGAlaLeu 819

RESULT 12
ABG61610
ID ABG61610 standard; Protein; 819 AA.
XX
AC ABG61610;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-2 splice variant #9.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskenesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN MO2002J1134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001MO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI O1 S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR MPI: 2002-444178/47.
XX
PT N-PSDB; ABRK3341.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
XX
PS Disclosure: Page 101-103; 113pp: English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskenesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 819 AA:

Alignment Scores:

Pred. No.:	0	Length:	819
Score:	4172.00	Matches:	789
Percent Similarity:	92.61%	Conservative:	0
Best Local Similarity:	92.61%	Mismatches:	1
Query Match:	87.01%	Indels:	62
DB:	23	Gaps:	2

US-09-976-674-4 (1-2617) x ABG61610 (1-819)

QY 11 ATGGCCACCAACCGGACCCCAAGCGCGACGAGCGAGCGACCGCCGACAGATGACCCG 70
|||||

DB 30 MetaLaThrThrGlyThrProThrAlaAspArgGLyAspAlaLaLaThrAspAspPro 49

QY 71 GCCGCCGCTTCAGAGTGCAGAAACACTCTGGGACGGCTCCGAGCATCATCAGCGC 130
|||||
DB 50 AlaAlaArgPheGlnValGlnIlyshISerTrpAspGLyLeuArgSerIleIleHisGLy 69
|||||
QY 131 AGCCGCAAGTACTCGGGCTCATTTGTCACAAAGGCGCCACAGATTCCAGTTTGTCAG 190
|||||
DB 70 SerArgLysTyrSerGlyLeuIleValIshnLysAlaProHisAspPheGlnPheValGln 89
|||||
QY 191 AAGACGATAGTCTGGGCCCCCACCCTCCAGCCGCTCTACTACCTCGGGAATGCCATAGCC 250
|||||
DB 90 LysThrAspLysSerGlyProHisSerHisArgLeuTyrTrpLeuGlyMetProTyrGly 109
|||||
QY 251 AGCCGAGAGAACTCCCTCTACTGTGAGATTCCCAAGAAAGTCCGGAAGAGGCTCTG 310
|||||
DB 110 SerArgLysnSerLeuLeuTyrSerGlnIleProLysLysValArgLysGlnAlaLeu 129
|||||
QY 311 CTGCTCTGTCTGGAAGCAAGATGCTGGATCTTCCAGGCCACGCCCATGGGCTC 370
|||||
DB 130 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisIShISGLyAl 149
|||||
QY 371 TACTCTGGGAGGAGAGGAGCTGTGAGGAGCGGAAGCCCTGGGGCTTCGGCATACC 430
|||||
DB 150 TyrSerArgLysGlnGlnLeuLeuArgGlnArgLysArgLeuGlyValPheGlyIleThr 169
|||||
QY 431 TCCCTACGACTTCCACAGCGAGAGTGGCTCTCTCTCCAGGCGACCAAGCGCTCTTC 490
|||||
DB 170 SerTrpAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
|||||
QY 491 CACTGTGCCAGCGGCGCAAGAACGGCTTCATGGTGTCCCTATGAAGCCGTGGAATC 550
|||||
DB 190 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 209
|||||
QY 551 AAGACCGAGTGTCTGAGGCGCCCGGATGAGCCCAAAATCTGCCCTGGACCCCTGCTTC 610
|||||
DB 210 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
|||||
QY 611 TTCTCCTTCATCATTAACAGCGACTGTGGTGCGCAACATCGAGACGCGAGAGCGG 670
|||||
DB 230 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGlnIleGlnIleArg 249
|||||
QY 671 CGGCTGACCTTGTGCCACCAAGGTTTATCCAAATGCTCTGATGACCCCAAGTGTGGGGT 730
|||||
DB 250 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 269
|||||
QY 731 GTGGCCACCTTCGTCATACAGGAAGATGTGACCGCTTCCTGGGTCTGGTGGTGGCCC 790
|||||
DB 270 ValAlaThrPheValIleGlnGlnIlePheAspArgPheThrGlyTyrTrpTrpCysPro 289
|||||
QY 791 ACAGCCTCTGAGGAAGGTTGAGAGGGCTCAAGACGCTGGCAATCCTGTATGAGGAATC 850
|||||
DB 290 ThrAlaSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnGlnVal 309
|||||
QY 851 GATGAGTCCAGGTGAGGAGTCATTACGTCCTCTCTGCTGCTAGAGAAAGAGACG 910
|||||
DB 310 AspGlnSerGlnValGlnValIleHisValProSerProAlaLeuGlnGlnArgLysThr 329
|||||
QY 911 GACTCGTATCGGTACCCAGGACAGCGAGCAAGAAATCCCAAGATTGCTTGAATCGGCT 970
|||||
DB 330 AspSerTrpArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
|||||
QY 971 GAGTTCAGACTACAGCGCGAGGACAGATGCTTCGACCCAGAGAAAGAGAGTGTGTCAG 1030
|||||
DB 350 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnLysGlnLeuValGln 369
|||||
QY 1031 CCCTTACGCTGCTGTTCGCCGAAGGTGAGTACATCGCCACGAGCGGGGTGACCCGGGAT 1090
|||||
DB 370 ProPheSerSerLeuPheProLysValGlnTyrIleAlaArgAlaGlyTrpThrArgAsp 389
|||||
QY 1091 GCGAATATACCCCTGGGCGCATGTTCTGTGAGACCGGCCACGAGATGGCTCCAGCTGCTC 1150
|||||
DB 390 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnIleThrLeuGlnLeuValLeu 409
|||||

QY 1151 CTCCTCCGCGGCTGTTTCATCCGAGCAGAGATGAGAGACGGCTGACCTGCC 1210
 |||||||
 Db LeupProAlaLeuPheIleProSerThrLysnGlnGlnGlnAlaSerAla 429
 QY 1211 AGAGCTGTCCCGAGAGATGTCACCGCTATGTGTTACGAGAGAGTCACCAAGCTGCG 1270
 |||||||
 Db ArgAlaValProArgAsnValGlnProTyrAlaValTyrGlnGlnValThrAsnValTrp 449
 QY 1271 ATCAATGTTCAATGACATCTTTATCCCTCCCTCCCAATCAGAGGAGAGAGAGAGCTGCG 1330
 |||||||
 Db IleAsnValHisAspIlePheTyrProPheProGlnSerGlnGlnLysAspGlnLeuLys 469
 QY 1331 TTTCTCCGCGGCGCAATGATGAGACCGGCTTGCCATTTGTCAAAAGTACCGCGCTT 1390
 |||||||
 Db PheLeuAlaGlnAsnGlnLysLysThrGlnPheCysHisLeuTyrLysValThrAlaVal 489
 QY 1391 TTAATAATCCAGGCTACGATTTGAGAGTCAGCCCTTCACGCGCGGAGAGATGAATTTAAG 1450
 |||||||
 Db LeuLysSerGlnGlnTyrAspTrpSerGlnProPheSerProGlnGlnLysAspGlnPheLys 509
 QY 1451 TGCCCATTAAGAAGAGATTTGCTGTACACAGCGGTGAATGGAGGTTTGGCGAGGAC 1510
 |||||||
 Db CysProIleLysGlnGlnIleAlaLeuThrSerGlnTyrGlnValIleuAlaArgHis 529
 QY 1511 GGCTCCAAAGATCTGGGTCAATGAGAGACCAAGGTGTACTTCCAGGCGCAGCAGGAC 1570
 |||||||
 Db GlySerLys-----GlyThrLysAsp 536
 QY 1571 AGCGCGCTGAGACACACACTTACGTGAGTACATAGAGCGCGCGGAGATGCTGACGC 1630
 |||||||
 Db ThrProLeuGlnHisHisLeuTyrValValSerTyrGlnIleAlaGlnIleValArg 556
 QY 1631 CTCACACAGCGCGCTTCTCCATAGCTGCTCCATGAGCCAGAACTTGACATGTTGCTC 1690
 |||||||
 Db LeuThrThrProGlnPheSerHisSerCysSerMetSerGlnAsnPheAspPheVal 576
 QY 1691 AGCCACTACACAGGCTGAGACAGCGCGCTGCGGACGCTTACAACTGAGCGGCGCC 1750
 |||||||
 Db SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerIlePro 596
 QY 1751 GAGCAGCAGCCCTGACACAGACAGCCCGCTTGTGGCTACATGATGAGGACGACGAC 1810
 |||||||
 Db AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGlnAlaAlaSer 616
 QY 1811 TGCCCCCGGATATGTTCTTCCAGAGATCTTCCATTTCCACAGCGCTCGAGTGTGCG 1870
 |||||||
 Db CysProProAspTyrValProProGlnIlePheHisPheHisThrArgSerAspValArg 636
 QY 1871 CTCTACGGCATGATCTTACAGCCCGCAGCGCTTGCAGCGAGGAGAGACACCCCGCTC 1930
 |||||||
 Db LeuTyrGlnMetIleTyrLysProHisAlaLeuGlnProGlnLysLysHisProThrVal 656
 QY 1931 CTCTTTGTATATGAGGCCCGCCAGGTGACGTGATTAATCTCTTCAAAAGCATCAAG 1990
 |||||||
 Db LeuPheValTyrGlnGlnProGlnValGlnLeuValAsnAsnSerPheLysGlnLys 676
 QY 1991 TACTTGGCGCTCAACACACTGCGCTCCCTGGGCTACAGCGCTGTGTATGACGCGCAGG 2050
 |||||||
 Db TyrLeuAlaGlnLeuAsnThrLeuAlaSerLeuGlnTyrAlaValValIleAspLysArg 696
 QY 2051 GGCTCCCTGTCAGGAGGCTTCCGTTGAGAGGCGCTGAAAGAACCAATGGCGCAGGCTG 2110
 |||||||
 Db GlySerCysGlnArgGlnLeuArgPheGlnGlnGlnAlaLeuLysAsnGlnMetLysIleVal 716
 QY 2111 GAGATCGAGGACAGAGTGGAGGCGCTGACATTCGTGGCCGAGAAATATAGCTTCTCAGC 2170
 |||||||
 Db GlnIleGlnAspGlnValGlnGlnLysLeuGlnPheValAlaGlnLysTyrGlnPheLeuAsp 736
 QY 2171 CTGACCGAGTTGCCATCATGCTGTGTCCTACGGGCGCTTCTGCTCATGCGGCGCTA 2230
 |||||||
 Db LeuSerArgValAlaIleHisGlnTyrPheTyrGlnGlnPheLeuSerLeuMetGlnLysLeu 756

QY 2231 ATCCAAAGCCCGAGGTTCATAGAGTGGCCATCCGCGGTGCCCGGTCACCGCTGATG 2290
 |||||||
 Db IleHisLysProGlnValPheLys--Ala----- 765
 QY 2291 GCCTACGACACAGGGTACACTGAGCGCTACATGAGACTCCTGTGAGAACACACGACGCG 2350
 |||||||
 Db 765 ----- 765
 QY 2351 TATGAGCGGGTTCGCTGCGCCCTGACGTGGAGAAAGCTGCCAATGAGCCCAACCGCTG 2410
 |||||||
 Db 766 -----GlnProLeuA 769
 QY 2411 CTATACCTCCAGGCTTCTGAGCAAAAGCTGACTTTTCCACAAACTTCTGCTC 2470
 |||||||
 Db IeTyrProProArgLeuProGlnTyrArgLysArgAlaLeuPheProHisLysLeuProArgL 789
 QY 2471 TCCCAACTGATTCGAGCAGGAGGAAACCTTACAGAGCTCCAGATCTACCCCAAGAGACAC 2530
 |||||||
 Db eupProThrAspProSerArgLThrLeuProAlaProAspLeuProGlnArgGlnThrG 809
 QY 2531 AGTATTCGCTGCGCGAGTGGGCGAGCAGCTA 2562
 |||||||
 Db InTyrSerLeuProArgValGlnArgAlaLeu 819
 RESULT 13
 AAE23875
 ID AAE23875 standard; Protein: 847 AA.
 AC AAE23875;
 AC AAE23875;
 DT 23-SEP-2002 (first entry)
 DE XX
 DE Murine dipeptidyl peptidase 9 (DPP9) protein.
 KW XX
 KW Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
 KW human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
 KW type II diabetes; antidiabetic; antiinflammatory; immunosuppressive;
 KW antiviral; enzyme.
 OS XX
 OS Mus sp.
 PN XX
 PN W0200234900-A1.
 PD XX
 PD 02-MAY-2002.
 PE XX
 PE 29-OCT-2001; 2001WO-AU01388.
 PR XX
 PR 27-OCT-2000; 2000AU-0001078.
 PA XX
 PA (UNSY) UNIV SYDNEY.
 PI XX
 PI Abbott CA, Gorell MD;
 PS XX
 PS WPI: 2002-454646/48.
 PS N-PSDB: AAD38311.
 PT XX
 PT New dipeptidyl peptidase (DPP) peptides, useful for screening
 PT inhibitors of DPP catalytic activity, which may be employed to treat
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 PT rejection and HIV infection -
 XX
 PS Claim 1; Fig 8; 91pp; English.
 CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
 CC polynucleotides encoding such proteins. The DPP peptides are useful for
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 CC rejection and HIV (human immuno deficiency virus) infection. The present
 CC sequence is an alternative version of murine DPP9 protein.
 CC Note: This sequence is stated to be the same as that shown as
 CC SEQ ID NO: 4 in the sequence listing of the specification. However these
 CC sequences differ.

SQ Sequence 847 AA:

Alignment Scores:

Pred. No.:	1.69e-313	Length:	847
------------	-----------	---------	-----

Score:	4129.00	Matches:	764
--------	---------	----------	-----

Percent similarity:	95.79%	Conservative:	32
Best local similarity:	91.94%	Mismatches:	35

```

2000 local similarity: 24.24% mismatches: 33
Query Match: 86.11% Indels: 0

```

```
DB:      23
Gaps:    0
```

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

US-09-976-b/4-4 (1-2617) x AAE23875 (1-847)

[illegible]

Db	310	argtgyrrroalgythnglyserlyasnpolylllealileuylsleuylaeliuleucln	111
Oy	980	actgcacaccagggcagaatcgtctgcagccacagagaaagacgtggtgcagcccttcacg	10339
Db	330	thraspnsinglylsylevalisersecyugilyuglyuvalglinproheser	349
Oy	1040	tcggcggtcccggaagggagatcatccagggccgggtggacccggatggcacaatac	10999
Db	350	serleupherpolyvalglyutyllealatalaglyltyrphtrargaspelylustytr	369
Oy	1100	gccttggcccatgttcctgcagaccggcccccacacatggcctccacotgctctctcccccgc	11599
Db	370	alatrralmetrheleuasrparcproglnlarginyleuclnleuvalleuupropro	389
Oy	1160	gccctgttcattcccgagacacagaaatgagagacacggctagacctctgcacagctgtc	12199
Db	390	alaleupherleerolalavalglusergualaglnatrglnalalatalargalaval	409
Oy	1220	cccgagaaattccacagccgtatgttgcttgcagagagctacacacagctgcgataatgtt	12799
Db	410	prolysaenvalglinprophelvaliletyuglyuglyvalthrasnvaltrprieasnval	429
Oy	1280	catgcacatctttatccctttcccccatacagagagagacacagctctgtttctccgc	13399
Db	430	hlsasrleupherpnspropherproglinalaglyuglylnlasphecyshleuarg	449
Oy	1340	gccatgaatgcacagaccggcttgcgcatgtttacaaagtcacccggctttaaatcc	13999
Db	450	alaasnuglyslsyttrnglyrhecyshlsleutyfargvalthryalcluleuylsthr	469
Oy	1400	cagggctcgaattgagtgagacccttcacggccggggaagatgaatttaagtcgcccaatt	14599
Db	470	lysasrtylasrprlnrlnrlnrleuaserprothnglyuglyuhelyselysrotle	489
Oy	1460	laagcagaagattgctctgcaccacgggtgaatggagacgtttggccagacacggcttcacg	15199
Db	490	lysgluginvalalaleuthtrsergluglytrglnvalleuserfarghslselys	509
Oy	1520	atctcgagcatgtgagagaccacagctgcttacttcacagggacacaaagacacggcgtg	15799
Db	510	iletrpvalasnuglylnrlylsleuvaltylrheclnglythlylasrphrproleu	529
Oy	1580	gagcacacacsttgcagctgctgcacatgcagacggccggcgagacgtgcacgcttcacacg	16399
Db	530	glunhlsileuylvalylaserlyuglyuserfalglylnlevalargheuthrthr	549
Oy	1640	ccggcgcttccccaactacgtctccacagacacaaattgcacatgctgtgcagccacac	16999
Db	550	leuglyrheesrthlssecyusermetserglnserphasrpherhelaserthistr	569
Oy	1700	agcacagctgacacacggccctcgctgcacagcttcacacagctgcagacggcccgacacac	17599
Db	570	sersevalserthrproprocysvalinsvaltylylseuuserglyproasrparasp	589
Oy	1760	ccccctgcacacagacggccgcttctggcgctagacatagtgagagacacacgctcccccgc	18199
Db	590	proleuhlslysglnproadurphertrpalaSermetmetglnalalasnycsprofo	609
Oy	1820	gattatgcttccctcacagagatcttcacatttccacacacggctcgagatgcggcttcacgc	18799
Db	610	asrpyrvalrproroglylilerphenlsrphenlsrthargalasalvalglinleuylcgly	629
Oy	1880	atgatctacacagcccccacggcttgacacacagagagacacacccacacggctctctttgta	19399
Db	630	metleutylysprohlsrthtrleuglnrprogluatrglyshlsrprothryalileupherval	649
Oy	1940	tatgcagacggcccccagcagctgcagctggagatgaattacttccttaaaaggatcaagactgcgg	19999
Db	650	tyrcltyglyproglinalglinleuvalasnasrserphelyscllyleusttytleuarg	669
Oy	2000	ctcaacacacactggcctcttgggctgcagccgctgtgtgattgcagcgacagggcctctgt	20599

Db 670 LeuAsnThrLeuAlaSerLeuGlyTyrAlaValAlaValIleAspGlyArgGlySerCys 689
 QY 2060 CAGGAGGGCTGGTGGTTCAGAGGGGCGCTGAAAAACCAATGAGCCAGCTAGATCCAG 2119
 Db 690 GlnTrpGlyLeuHisPheGlnGlyAlaLeuAlaSerGlnMetGlyGlnValGlnIleGln 709
 QY 2120 GACCAAGTGAAGGCTTCAGTTCGTGGCGCAGAAAGTATGGCTTCATGACCTGAGCCGA 2179
 Db 710 AspGlnValGlnGlyLeuGlnGlyTyrValAlaGlnIlyTyrGlnPheIleAspLeuSerArg 729
 QY 2180 GTTCCCATCCATGAGCTGCTCTACGGGGCTTCCTCTGCTCATGAGGGCTATCCACAAG 2239
 Db 730 ValAlaIleHisGlyTyrPheSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLys 749
 QY 2240 CCCAGGTGTTCAGTGTGCTCCATCGCGGGTCCCGGCTGACCCGTCTGATGGCTACAGAC 2299
 Db 750 ProGlnValPheLysValAlaIleAlaIleAlaProValIleTyrPheAlaIlyrAsp 769
 QY 2300 ACAGGGTACACTGAGCCGTATCATGAGCTCCCTGAGAACACAGCAGCGCTATGAGCG 2359
 Db 770 ThrGlyTyrThrGlnIlyrGlyTyrMetAspValProGlnAsnGlnGlnGlyTyrGlnAla 789
 QY 2360 GGTTCGGTGGCCCTGCAGCTGAGAGAGCTGCCAATGAGCCCAACCGCTTCTTATCCTC 2419
 Db 790 GlySerValAlaLeuHisValGlnIlyLeuProAsnGlnIlyProAsnArgLeuIleuIleu 809
 QY 2420 CAGGGCTTCCTGGAGAGAAAGCTGACCTTTTCCACACAACCTTCTGCTCCCACTG 2479
 Db 810 HisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuValSerGlnIleu 829
 QY 2480 ATCCGAGCAGGAGAACTTACCATGCTCCAGTCCAGATC 2512
 Db 830 IleArgAlaGlyLysProTyrGlnLeuGlnIleVal 840
 RESULT 14
 AAM38724
 ID AAM38724 standard; Protein: 737 AA.
 AC AAM38724:
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 1869. *no snrln seq.*
 XX
 KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HXSE-) HXSEQ INC.
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX
 DR WPI: 2001-442253/47.
 DR N-PSDB: AA157880.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 3: SEQ ID NO 1869; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 SO Sequence 737 AA:
 Alignment Scores:
 Pred. No.: 3,3e-289 Length: 737
 Score: 3817.50 Matches: 721
 Percent Similarity: 94.64% Conservative: 3
 Best Local Similarity: 94.25% Mismatches: 4
 Query Match: 79.61% Indels: 37
 Dbs: 22 Gaps: 4
 US-09-976-674-4 (1-2617) x AAM38724 (1-737)
 QY 332 ATGCTGATCATTTTCACAGCCAGCCGACCATGAGGGGTCTACTCTGGAGAGAGAGCTG 391
 Db 1 MetLeuAspHisPheGlnAlaThrProHisHisGlyValTyrSerArgGlnGlnIleu 20
 QY 392 CTGAGGAGCGGAGAAAGCGCTGGGGGCTTTCGGCATCACCTCTAGACTTCCACAGCGAG 451
 Db 21 LeuArgGlnAlaGlyLysArgLeuGlyValPheGlyIleThrSerTyrAspPheHisSerIleu 40
 QY 452 AGTGAGCCTCTTCCTTCAGAGCGCAGCAAGACCTTCCACTGCGGAGCGGCGAGAG 511
 Db 41 SerGlyLeuPheLeuPheGlnAlaSerHisSerLeuPheArgCysArgAspGlyGlyLys 60
 QY 512 AACGGCTTCATGCTGCTCCCTATGAAACCGCTGAATCAAGACCAAGCTCTCAGGGGCC 571
 Db 61 AsnGlyPheMetValSerProMetLysProLeuGlnIleLysThrGlnCysSerGlyPro 80
 QY 572 CGGATGAGACCCCAAAATGCGCCCTGCGACCTGCTCTCTCTCATCAATAAGACG 631
 Db 81 ArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPheIleAsnSer 100
 QY 632 GACCTGTGGTGGCCACATCGAGACAGAGAGAGCGGCGCTGACCTTGTGCACCAAA 691
 Db 101 AspLeuTrpValAlaAsnIleGlnThrGlnGlnAlaGlnGlnThrPheCysHisGln 120
 QY 692 GGTTCATCCATGCTCTGAGTACGCCCAAGTCTGCGGGTGTGGCCACCTTGCATACAG 751
 Db 121 GlyLeuSerAsnValLeuAspAspProLysSerAlaGlyValAlaThrPheValIleGln 140
 QY 752 GAAGAGTTCGACCGCTTACAGGGTACGAGTGTGCTGCCCAAGCGCTCTGGAAGAGTCA 811
 Db 141 GlnGlnPheAspArgPheThrGlyTyrTrpTrpCysProThrAlaSerTrpGlnIlySer 160
 QY 812 GAGGCGCTCAAGACCGCTCGAATCCTGTATGAGAGAGTGCATGATCCGAGGTGAGGCTC 871
 Db 161 GlnGlyLeuLysThrLeuArgIleLeuTyrGlnGlnValAspGlnSerGlnValGlnVal 180

QY	872	ATTCAGCTCCCTCTCTCTCGGGCTTAGAAGAAAGAAACAGACGACTCGATGTGGTACCCACAG	931
Db	181	IleHisValProSerProIalaLeuGluGluValGlyThrAspSerTyrArgTyrProArg	200
QY	932	ACAGGACACAGAAATCCCAAGATGGCTTAAACTGGGTGAGTTCACAGCTACAGCCAG	991
Db	201	ThrGlySerLysAsnProLysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGln	220
QY	992	GGCAAGATCGTCTGCAGCCACAGAGAAAGAGACTGTGCAGCCCTCAGCTCGCTGTTCGG	1051
Db	221	GlyLysIleValSerThrGlnGluLysGluLeuValGlnProPheSerSerLeuPhePro	240
QY	1052	AAGGGGAGTCAATCGCCACAGGGCGGGTGGACCCGGGATGGCAAAATAGCCCTGGGCCAAG	1111
Db	241	LysValGluTyrIleAlaAlaArgAlaGlyTyrThrArgAspGlyLysTyrAlaTrpAlaMet	260
QY	1112	TTTCCTGACCGGGCCACAGAGTGGCTTCACAGCTGCCTCTCTCCCGCCGCTGTTCATC	1171
Db	261	PheIleuAspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProIalaLeuPheIle	280
QY	1172	CCGAGCACAGAAATGAGAGAGACAGCGGCTTAGCTCTGCCAGAGCTGTCCCAAGAAATGC	1231
Db	281	ProSerThrGluAsnGluGluGlnArgLeuAlaSerAlaArgAlaValProArgAsnVal	300
QY	1232	CAGCGGTGTGGGTACGAGAGAGAGTACCACAGCTGTGGATCAATGTTATCATATCTTC	1291
Db	301	GlnProTyrValValTyrGluGluValAlaThrAsnValTrpIleAsnValHisAspIlePhe	320
QY	1292	TATCCCTTCCCCCAATCAGAGGAGAGAGACAGACTGTGTTCTCTCCGCCCAATGAATGC	1351
Db	321	TyrProPheProGlnSerGluGluGlyLysAspGluLeuCysPheLeuAlaArgAlaAsnGluCys	340
QY	1352	AAGACCGGCTTCGCCATTGTGTCAAAAGTACCGCGGTTTAAATCCAGGGCTACAGAT	1411
Db	341	LysThrGlyPheCysHisLeuTyrLysValAlaValAlaLeuLysSerGlnGlyArgTSP	360
QY	1412	TGGAGTGACCCCTTCAGCCCCGGGAGAAAGATGAATTTAAGTGGCCCATTAAGAAAGAGATP	1471
Db	361	TrpSerGlnProPheSerProGlyGluGly-----GluGln 372	
QY	1472	GCTCTGACACAGGGTGAAATGGAGGTTTGGCAGGGCAGGGCTCAAGATCTGGGCTCAT	1531
Db	373	SerLeuThrAsnAla-----IleTrpValAsn 381	
QY	1532	GAGAGACACAGACTGTGTACTTCCAGAGGCCACCAAGACAGCGCGGTGGAGCCACCACTC	1591
Db	382	GluGluThrLysLeuValTyrPheGlnGlyThrLysAspThrProLeuGlnHisHisLeu	401
QY	1592	TACGTGGTCAAGTATGAGGGGGCGGGCGAGATGTACGCGCTACACAGCGCGGCTTCTCC	1651
Db	402	TyrValIleSerTyrGluAlaAlaGlyGluIleValAlaArgLeuThrThrProLysPheSer	421
QY	1652	CATAGCTGCTCCATGAGACCCAGAACTTCGACATGTTGGTACGACACTACAGCGAGCTGTAGC	1711
Db	422	HisSerLysSerMetSerGlnAsnProPheAspMetPheValSerHisTyrSerSerValSer	441
QY	1712	ACGCGCGCTTGCATGACGTCTACAAGCTGAGCGGGCCGACAGACGACGCCCTGTGCACAG	1771
Db	442	ThrProProCysValHisValTyrLysLeuSerGlyProAspArgAspProLeuHisLys	461
QY	1772	CAGCGCGGCTTCGGGCTACCATGATGAGAGCGACGCTGCCCGCCGATTAATGTCTCT	1831
Db	462	GlnProArgPheTrpAlaSerMetMetGluAlaAla----- 473	
QY	1832	CCAGAGATCTTCATTTCCACAGCGGCTGGATGTGCGGCTCTACGAGCATGATCTACAG	1891
Db	474	---LysIlePheHisPheHisThrArgSerAspValAlaArgLeuTyrGlyMetLeuTyrLys	492
QY	1892	CCCCACGCTTCGACGCGAAGAAAGAACACCCACGCTCTCTTTGTATATGAGAGCCCC	1951
Db	493	ProHisAlaLeuGlnProGlyLysHisProThrValLeuPheValTyrGlyPro 512	

QY	1952	CAGTGGACGCGTGGTAAATACCTCTTAAAGGATGACAAAGACTTGTGGGGCTCAACACACTG	201
Db	513	GTnValGlnLeuValAsnAsnSerPheLeuGlyIleLeuSTyTLeuArgLeuAsnThrLeu	532
QY	2012	GGCTCCCTGGGGCTACGGCGTGGTGTATTATGACGGGACGGGGGCGCTGTACGGAGGGGCTT	2071
Db	533	AlaSerLeuGlyTyrIleAlaValIleAlaSerPheArgGlyIleArgGlySerCysGlnArgGlyLeu	552
QY	2072	CGGTTTCAAGGGGGCCCTGAAAAACAATGGCCAGGTGAGATGTGACAGACCAAGGTGGAG	2133
Db	553	ArgPheGlnGlnIleAlaLeuLysAsnGlnMetGlyGlnIleGlnIleGlnAspGlnIleAlaGln	572
QY	2132	GAGCGTCAGTTGCTGGCCGAGAAATATGGCTTATCATGACCTGAGCCGAGTTCCATTCAT	2191
Db	573	GlyLeuGlnPhePheValAlaGlnLysTyTGlyPheIleAspLeuSerArgValAlaIleHis	592
QY	2192	GAGTGGTCTTACGGGGGCTCTCTCGCTCATGGGGCTAATGCSACAAAGCCCAAGGTGTTC	2251
Db	593	GlyTyrSerTyrGlyIlePheLeuSerLeuMetGlyLeuIleHisLysProGlnIleAlaPhe	612
QY	2252	AAGGGGCCAATGGGGGGTGGCCCGGTACCGGTGTGGATGGGCTACGACACAGGGTAACT	2311
Db	613	LysValAlaIleAlaGlyAlaProValThrValTyrMetAlaTyrAspThrIleTyrThr	632
QY	2312	GAGCGCTCATGAGACGTCCCTGAGAACACACACAGCGCTATGAGCGGGTTCGGTGGCC	2371
Db	633	GlnArgTyrMetAspValProGlnAsnAsnGlnHisGlyTyrGlnIleAlaIleSerValAla	652
QY	2372	CTGCACGTGGAGAAAGTCCCAATAGACCCACACCGCTGTGCTTATCTCCACAGGCTTCTG	2431
Db	653	LeuHisValGlnLysLeuProAsnGlnProAsnArgLeuLeuIleLeuHisGlyPheLeu	672
QY	2432	GACGAAACGTCGACTTTTTCACACAAATCTCTGTCTCCCACTGATTCGAGACAGG	2491
Db	673	AspGlnAsnValHisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGly	692
QY	2492	AAACCTTACGAGTCTC-----CAGATCTAACCCCAACGAG	2524
Db	693	LysProTyrGlnLeuGlnValAlaLeuProProValSerProGlnIleTyrProAsnGln	712
QY	2525	AGACCACTATTTCGCTGCCCGAGTGGGGGAGCACTATGAAGTCACTGTGGACTTTT	2584
Db	713	ArgHisSerIleArgCysProGlnSerGlyGlnHisTyGlnValThrLeuLeuHisPhe	732
QY	2585	CTACAGAAATACCTC	2599
Db	733	LeuGlnGlnTyrLeu	737
RESULT 15			
AA040510			
ID	AA040510 standard; Protein; 683 AA.		
XX			
AC	AA040510;		
XX			
DDT	22-OCT-2001 (first entry)		
XX			
DE	Human polypeptide SEQ ID NO 5441.		
XX			
KW	Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
OS	Homo sapiens.		
XX			
PN	MO20015312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000MO-US34263.		

PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI59666.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2: SEQ ID NO 5441: 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA438642-AA442213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 SQ Sequence 683 AA;
 Alignment Scores:
 Pred. No.: 1.62e-267 Length: 683
 Score: 3539.50 Matches: 668
 Percent Similarity: 94.37% Conservative: 3
 Best Local Similarity: 93.95% Mismatches: 37
 Query Match: 73.82% Gaps: 4
 DB: 22
 US-09-976-674-4 (1-2617) x AAMA0510 (1-683)
 OY 494 TGTGCGACGGCGGCAAGACGGCTTCATGCTGCCCTATGAACCGCTGGAATCAAG 553
 Db 1 CysAtgAspArgIlyGlyAsnGlyPheMetValSerPrometIlysProLeuGluIleIys 20
 OY 554 ACCGAGTCTAGAGCGCCCGGATGAGACCCCAAAATCTGCCCTGCCGACCTGCTTCTC 613
 Db 21 ThrIlnCySerGlyProAlaMetAspProIlySileCySProAlaAspProAlaPhe 40
 OY 614 TCTTCATCATTAACAGCGACCTGTGAGTGGCCAAACATCGAGACAGGAGGAGCGGCG 673
 Db 41 SerHeIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluAlaArg 60
 OY 674 CTGACCTTCTGCCACCAAGTTTATCAATGTCTGATGACCCCAAGTCTGCGGTGTG 733
 Db 61 LeuThrPheCyShISnGlnGlyLeuSerAsnValIleuAspAspProIlySerAlaIylal 80
 OY 734 GCGACCTTCTGCATACAGAGAGTTCGACCGCTTCACCTGAGTACTGGTGGTGGCCCA 793
 Db 81 AlaThrPheValIleGlnGlnIleuAspArgPheThrGlyTyrTrpTrpCysProThr 100
 OY 794 GCCTCTCTGGAGGTTTCAGAGGCGCTCAAGACGCTCGAATCCTGTATGAGGAAGTGCAT 853

Db 101 AlaSerTrpGlnGlySerGlnGlyLeuIlySerThrLeuArgIleLeuTyrGlnGluValAsp 120
 OY 854 GAGTCCGAGGTGAGAGGTATTCACGTCCCTCTCTGCGCTAGAGAAAGAGACAGCGAC 913
 Db 121 GluSerGlnValAlGlnValIleHisValProSerProAlaLeuGlnGluArgIlyThrAsp 140
 OY 914 TCGTATCGGTACCCGAGGACGAGGACGAAGATCCCAAGATGGCTTGAACTGGCTGAG 973
 Db 141 SerTyrArgIlyTrpAlaGlyThrGlySerIlySAsnProIlySileAlaLeuIlySleuAlaIu 160
 OY 974 TTCAGACTGACACGACGAGGCAAGATGCTGTGACCCGAGAGAGAGAGTGTGTCAGCCC 1033
 Db 161 PheGlnThrAspSerGlnGlyIleValSerThrGlnGlnIlySleuValGlnPro 180
 OY 1034 TTCAGCTCGCTGTTCGAGAGGTGAGTACATCCGACGCGCGGTGAGACCCGGATGCG 1093
 Db 181 PheSerSerLeuPheProIlyValGlnIlyTrpIleAlaArgAlaGlyTrpThrArgAspIly 200
 OY 1094 AAATACGCGCTGGCGCATGTCTGTGACCGCGCCGAGCGATGCTGCTGCTCTC 1153
 Db 201 LysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnIlyPheGlnIlyLeuValLeu 220
 OY 1154 CCCCCGCGCTGTTCATCCCGAGGACACAGAGATGAGAGACGCGCTAGCCTGTGCGAGA 1213
 Db 221 ProProAlaLeuPheIleProSerThrGlnGlnGlnGlnIlyAlaSerAlaArg 240
 OY 1214 GCTGTCCCGAGGAATGTCACCGCTATGTGTTGACAGAGAGTCCACACGCTGTGATC 1273
 Db 241 AlaValProAlaGAsnValAlaProIlyValAlaTyrGlnGluValThrAsnValTrpIle 260
 OY 1274 AATGTATGACATCTTATCCCTTCCCAATACAGAGGAGAGGAGACGCTGTGCTT 1333
 Db 261 AsnValHisAspIlePheTyrProPheProGlnSerGlnGlyGluAspIlyLeuCysPhe 280
 OY 1334 CTCGCGCGCAATGATGACAGACCGCGCTTCTGCCATTGTGACAAAGTACCGCGCTTTA 1393
 Db 281 LeuArgAlaAsnGlnCysIlySerThrGlnIlyPheCysHisLeuTyrIlyValThrAlaValLeu 300
 OY 1394 AAATCCGAGGCTACGATGAGAGTACGACCTTCACCGCGGGAAGATTAATTAAGTGC 1453
 Db 301 LysSerGlnGlyTyrAspTrpSerGlnProPheSerProGlnGlyGly 316
 OY 1454 CCCATTAAGAAAGATTGCTGTGACGCGGTGAATGGAGGTTTGGCAGGACGCGC 1513
 Db 317 -----GluGlnSerLeuThrAsnAla----- 323
 OY 1514 TCGAAGATCTGGTTCATACGAGAGACGCAAGCTGTGACTTCCAGGCGACCAAGACAGC 1573
 Db 324 -----IleTrpValAsnGlnGlnIlyThrIlySleuValTyrPheGlnGlnIlyThrIlyAspTrp 341
 OY 1574 CCGCTGGAGACACACCTGTAGCTGAGTACAGTATAGAGCGCGCGGAGATGCTAGCCTGC 1633
 Db 342 ProIleuGlnHisIleLeuTyrValValSerTyrGlnAlaAlaGlyIleValArgLeu 361
 OY 1634 ACCAGCGCGCGCTCTCCATAGCTGTGCTGACGCGCAAGCTTGGATGTGCTGACG 1693
 Db 362 ThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheValSer 381
 OY 1694 CACTACAGCAGCGTGTGAGACGCGCGCTGTGCTGACGCTGACAGCTGACGAGCGCGCCAG 1753
 Db 382 HisTyrSerSerValSerThrProProCysValHisValTyrIlySleuSerGlyProAsp 401
 OY 1754 GACGACCGCTGTGCACAGACGCGCGCTTGTGAGCTACAGATGAGGAGGACCGACGCTGC 1813
 Db 402 AspAspProLeuHisIlyGlnProArgPheTrpAlaSerMetMetGlnAlaAla----- 419
 OY 1814 CCCCCGATTAATCTCCSCAGAGATCTTCATTCACAGCGGCTGGATGTGCGGCTC 1873
 Db 420 -----LysIlePheHisPheHisIlyThrArgSerAspValArgLeu 432
 OY 1874 TACGCGATGATCTACAGCCCGACGCTTGCAGCGGAGGAGAGACCGACCGCTCTC 1933

```
Db 433 TyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysHisProthValIleu 452
QY 1934 TTGTATATGAGAGCCCCAGAGTGCAGCTGTGTAATACCTTCAAGGATCAAGTAC 1993
Db 453 PheValTyrGlyGlyProGlnValGlnLeuValAsnSerPheLysGlyIleLysTyr 472
QY 1994 TTGCGGCTCAACACACTGAGCTCCCTGGGCTACGCCCTGTGATTGACGGCAGGGGC 2053
Db 473 LeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGly 492
QY 2054 TCCTGTACAGGAGGGCTTCGGTTCCGAAGGGCCCTGAAAAACCAATGGCCAGGTGAG 2113
Db 493 SerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnValGlu 512
QY 2114 ATCGAGGACAGGTGAGGGCTGCAGTTCGTGGCCGAGAAGTATGGCTTCATCGACCTG 2173
Db 513 IleGlnAspGlnValIleGlnGlyLeuGlnPheValAlaGlnLysTyrGlyPheIleAspLeu 532
QY 2174 AGCCGAGTTGCCATCCATGGCTGGTCCACGGGGCTTCCTCTGCTCATGGGGCTAATC 2233
Db 533 SerArgValAlaIleHisGlyTyrSerTyrGlyGlyPheLeuSerIleuMetGlyLeuIle 552
QY 2234 CACAAGCCCCAGGTGTCAAGGTGCGCATGCGGGGTGCCCCGGTCAACCTGTGATGGCC 2293
Db 553 HisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValIleTyrMetAla 572
QY 2294 TACGACACAGGATACACTGAGCGGTACATGACGTCCCTGAGAACACAGCAGCGCTAT 2353
Db 573 TyrAspThrGlyTyrThrGlnAlaGlyTyrMetAspValProGlnAsnAsnGlnHisGlyTyr 592
QY 2354 GAGGCGGGTCCGTGGCCCTGCACGTGGAGAGCTGCCCAATGAGCCCAACCGCTTGCTT 2413
Db 593 GlnAlaGlySerValAlaLeuHisValGlnLysLeuProAsnGlnProAsnArgLeuLeu 612
QY 2414 ATCTCCACAGGCTTCCTGGACGAAACAGTCACTTTTCCACACAACTCTCGTCTCC 2473
Db 613 IleLeuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuValSer 632
QY 2474 CAACTGATCCGAGCAGGGAACCTTACAGCTC----- 2506
Db 633 GlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnValAlaLeuProProValSerPro 652
QY 2507 CAGATCTACCCCAAGAGAGACACAGTATTCGCTGCCCGAGTCGGGCGAGCACTATGAA 2566
Db 653 GlnIleTyrProAsnGlnArgHisSerIleArgCysProGlnSerGlyGlnHisTyrGln 672
QY 2567 GTACAGTTGCTGCACCTTCTACAGGAATACCTC 2599
Db 673 ValThrLeuLeuHisPheLeuGlnGluTyrLeu 683
```

Search completed: December 12, 2002, 11:48:13
Job time : 106 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 12:04:17 : Search time 18 Seconds
(without alignments)
1410.666 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646
Sequence: 1 MATGTGPTADRGDAATDDP.....CPESGHEVTLHLFLEQYL 863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1093.5	23.5	310	US-09-794-236-4	Sequence 4, Appli
2	489.5	10.5	759	PCT-US93-07923-2	Sequence 2, Appli
3	489.5	10.5	766	US-08-230-491A-3	Sequence 3, Appli
4	489.5	10.5	766	US-08-619-280A-3	Sequence 3, Appli
5	489.5	10.5	766	US-08-940-391-3	Sequence 3, Appli
6	489.5	10.5	766	US-09-794-236-1	Sequence 1, Appli
7	482.5	10.4	755	PCT-US93-07923-3	Sequence 3, Appli
8	458	9.9	771	US-09-462-284-2	Sequence 2, Appli
9	426	9.2	760	US-08-230-491A-2	Sequence 2, Appli
10	426	9.2	760	US-08-619-280A-2	Sequence 2, Appli
11	426	9.2	760	US-08-940-391-2	Sequence 2, Appli
12	249	5.4	657	US-09-355-166-1	Sequence 1, Appli
13	240.5	5.2	632	US-09-016-080-1	Sequence 1, Appli
14	207	4.5	593	PCT-US93-07923-11	Sequence 11, Appli
15	163.5	3.5	721	US-09-390-234-20	Sequence 20, Appli
16	163.5	3.5	686	US-09-368-169-8	Sequence 8, Appli
17	160	3.4	622	US-08-664-646A-2	Sequence 2, Appli
18	160	3.4	622	US-09-066-285-2	Sequence 2, Appli
19	160	3.4	622	US-09-261-006-2	Sequence 2, Appli
20	160	3.4	622	US-08-951-088-2	Sequence 2, Appli
21	160	3.4	622	US-09-609-566-2	Sequence 2, Appli
22	160	3.4	622	US-09-609-570-2	Sequence 2, Appli
23	110.5	2.4	255	US-09-355-166-3	Sequence 3, Appli
24	107.5	2.3	346	US-08-602-359A-34	Sequence 34, Appli
25	107.5	2.3	422	US-08-485-938A-34	Sequence 34, Appli
26	106.5	2.3	444	US-08-485-938A-33	Sequence 33, Appli
27	103.5	2.2	614	US-08-446-100-21	Sequence 21, Appli

28	103.5	2.2	614	US-08-446-100-22	Sequence 22, Appli
29	103.5	2.2	614	US-08-446-100-23	Sequence 23, Appli
30	102	2.2	425	US-08-853-659A-38	Sequence 38, Appli
31	101	2.2	809	US-09-186-276B-58	Sequence 58, Appli
32	101	2.2	809	US-08-842-445-58	Sequence 58, Appli
33	101	2.2	809	US-09-186-188B-58	Sequence 58, Appli
34	100	2.2	1382	US-08-737-715-2	Sequence 2, Appli
35	100	2.2	1382	US-08-457-040B-7	Sequence 7, Appli
36	98	2.1	898	US-08-474-379C-12	Sequence 12, Appli
37	98	2.1	898	US-09-146-249A-12	Sequence 12, Appli
38	98	2.1	898	US-08-206-188B-12	Sequence 12, Appli
39	98	2.1	900	US-07-688-352C-12	Sequence 12, Appli
40	97	2.1	657	US-09-370-368-7	Sequence 7, Appli
41	96.5	2.1	870	US-08-785-241-4	Sequence 4, Appli
42	96.5	2.1	870	US-09-374-454-6	Sequence 6, Appli
43	96.5	2.1	963	US-08-537-002A-3	Sequence 3, Appli
44	96.5	2.1	963	US-08-863-010-3	Sequence 3, Appli
45	96.5	2.1	963	US-09-024-429-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-09-794-236-4
: Sequence 4, Application US/09794236
: Patent No. 6337069
: GENERAL INFORMATION:
: APPLICANT: Grouzmann, Eric
: APPLICANT: Lacroix, Jean-Silvain
: TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
: FILE REFERENCE: 81985/275823
: CURRENT FILING DATE: 2001-02-28
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: Patent version 3.0
: SEQ ID NO 4
: LENGTH: 310
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-794-236-4

Query Match      23.5%   Score 1093.5: DB 4: Length 310:
Best Local Similarity 57.3%: Pred. No. 1e+101:
Matches 200: Conservative 42: Mismatches 58: Indels 49: Gaps 1:
```

QY	515	FOGTRDFLEHNLVYVSYEAAGEIYRLTPGFSHSCSMSONFDMFVSHYSSVTPCYHV	574
DB	1	FEGTDSPLEHNLVYVSYVNPGEVTRLRDRGSHSCCSQHDPRISXNKNHCYSL	60
QY	575	YKLSGPPDDLPKOPRFNMAEMEAASCPDDVYPRPIFHFRSDVRLYGMITYKPHALOPG	634
DB	61	YKLSPPDDPCKTEKFATIDSAGPLDYPPIFFSESTGPTLYGMITYKPHALOPG	120
QY	635	KHNPVLVYGGPOQVIVNNSFKIKYLRNLALGVAVVYIDRGSCQRLREGALK	694
DB	121	KKYPVLVLYGGPO-----	134
QY	695	NOMGOVEIEDQEGLOAEKYGFIDLRSVAIHGMSGFSLMGLHKPOYFKVAIGA	754
DB	135	---GGIEIDDOVEGLQYLAIRDFIDLDRVGHGMSGYGLSLMALMROSDIFRAVAIGA	191
QY	755	PVTVMAYDTGTYTERRYMDVPPENNGHYEAGSYALHVEKLPENRLLIHGLDENVHF	814
DB	192	PVTLMIFDTGTYTERRYMDVPPENNGHYEAGSYALHVEKLPENRLLIHGLDENVHFA	251
QY	815	HNFVLSQLIRGKPYOLQIYENRHSIRCPESGHEVTLHLFLOEYL	863
DB	252	HNSILSLFVRAGKPYDQIYPOERHSIRVPSGHEVTLHLFLOEYL	300

RESULT 2

PCT-US93-07923-2
Sequence 2, Application PC/TUS9307923
GENERAL INFORMATION:
APPLICANT: Morimoto, Chikao
APPLICANT: Schlosman, Stuart F.
APPLICANT: Tanaka, Toshiaki
TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 759
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-2
Query Match 10.5%; Score 489.5; DB 5; Length 759;
Best Local Similarity 22.4%; Pred. No. 4.9e-40;
Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;
QY 47 LIYKAPHDFOFVOKTDSGPHSHRLYYLGMPIYSGRENSLLYSEIPKVKREALLLSWK 106
Db 19 VLKNGG-----TDDATFADSRRKTYTL-----TDYLNKTYRLKLYSLRW- 55
QY 107 QMDHFOATPHHGYRSREEL-----REKRLGVEGITSYDFH-SSSGLELF 153
Db 56 -----ISDEHYLYKQENNLVNAEYGNSSVLENSFDESHSINDSISIPDGGFIL 108
QY 154 QASNSLFHCRDGRKNGFVSPMKPLEIKTQCSGPRMDPKICPADPAF-FSFINNSDLAWA 212
Db 109 LEYVYVQWHRSTYASDIYDLNKRQLITEBRIINNNTQWYVSGHKLAVVWANDIYVK 168
QY 213 NIEGGERLLTFCHQIGLSNVLDPKSGAVTEVIQEE-FDRFTGYWMCPTASWEGSEGLK 271
Db 169 IEPNLPYRITW--TGKEDIIYN---GITDMVYEEVFSAYSLAMSPNGTF----- 215
QY 272 TLRLIVEVDESEYEVH---VSPALAEKRTDSYRPRGSKPKALKALAEQOTSOG 328
Db 216 ---LAAQFNDTEVPLLEYFSYDESILQYPRYKVPYKGAANPVY--KEFVYVNTSLS 270
QY 329 KIVSTOKELVOPFSSLFPRVEYIARAGWTRDGRYANAMEFLDRPQMLQVLPLPALFIP 388
Db 271 SVYNNATSIQITAPASMLIGD-HYLCDVYTAQTGERS-----LQML----- 309

QY 389 STENEORLASARAVPRNVQPVYV---YEEVTNWVINVHDIPEYPOSGEDELGLR 443
Db 310 -----RIQYVSWIDICDYDESSGRW----- 330
QY 444 ANECKTGCHLYKXTAVLAKSGYDMSEFSGDE-----FK 480
Db 331 --NCLVAROHEMSTT-----GVMGRFRSEPHFTLDGNSFYKIIISNEGRHICYFO 381
QY 481 CPICEIALTSGEWEVLARHSGKIMVNEETKLVEFGCKDPLRHHLYVSYEAG--- 536
Db 382 IDKDCFFITGTGTEVIGIAL-----TSD-----YLYISNEYKMGPG 421
QY 537 -EIVRLTTPGFSSCSMSQNDMEFVSHYSYSTPCVHVYKL--SGP----- 580
Db 422 RNLKYLQSDYTKVTCSCLEINPRCOYYSFSGKEAKYVQLRCSGGLPLYLTHSSVND 481
QY 581 -----DDDLHNOFPRFWSMMEASCPRPYVPELFHFTBDRVL-YMITYKPAHQ 632
Db 482 KGLRVEEDNSALDK-----MLQNVQ---MPSKLDFFILNETKFWYQMLPPLH-FD 528
QY 633 PGKHPVLYFYGGPOVQVLVNSFEKIKYLRN---TLASLGVAVVY-IDGRSGCQGLR 688
Db 529 KSKYVPLLDVYVYACPCQKADTF-----RLNMTATLASTENIIVASFDRSGSGYGDK 582
QY 689 FEGALKNMGQVEIEDQVEGL-QFVAEKYGFIDLSRVAIHGMSYSGELSLMGLIHKPOVF 747
Db 583 IMHAINRLGTFFVEEDQIEAARQF--SKMGFVDNKRRIAIMGMSYGVYVTSVLGSGGVF 640
QY 748 KVALAGAPVYVMAVYDGYPTFRYMDV--PENQNGYEGSVALVKEKLPNEPNLLILHG 805
Db 641 KCGIAPVAPVSRWEYDSDYTERYMGFLPPEEDNLDHRYNSTVMSRAENF--KQVEYLLIHG 698
QY 806 FLDENVHFHTNFLVYSLIRAKRKYOLQIYPNERHSIRCPESGHEVETLHFLFOE 861
Db 699 TADNVHFQDSQSLKRLVYGVDFQAMWITDEBHGIASSIAHQHITYHSHFIKQ 754
RESULT 3
US-08-230-491A-3
Sequence 3, Application US/08230491A
Patent No. 5587299
GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
APPLICANT: Garlin-Chesa, Pilar; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FELSE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-230-491A-3

```

Query Match	10.5%;	Score 489.5;	DB 1;	Length 766;
Best Local Similarity	22.4%;	Pred. No. 5e-40;		
Matches 201;	Conservative 124;	Mismatches 330;	Indels 241;	Gaps 35;

```

QY 47 LIIVKAPHDFOFQVOTKTESGHSRRLYLILMPYGSRENSLLYEIIPKVKKALLLSMK 106
Db 26 VLLKNG-----TDDATFADSKRYTL-----IDYKNTYRLKLYSRW- 62
QY 107 QMLDHFQATPHHGVYSREELL-----REKRRLGVFGTJSYDH-SESGLEF 153
Db 63 -----ISDHEYLYKOEENNLVFNAEYGNSSVPLENSTEDFERGHSINDYSISPDQOFL 115
QY 154 QASLSLHCRDGGKNGWMSVSMKLEIKTQCSGFRMDPKICSPDPAF-PSFINNSDLMA 212
Db 116 LEYVYVQWHRSHSYATSDIYDLNKRQOLITEERIPNNTQWTSVPVGHKLAYVNNNDIYK 175
QY 213 NIENGEBRRLFFHQGJLSNVLDDBKSAVATFIOE-PRFRPGYMMCSPASMEGSGLK 271
Db 176 IEPMLPYRITW--TGKEDIYN---GTDWYEEEFVFSAYLSALMWSPKGT------ 222
QY 272 TLRLIYEEVDESEVEVH---VPSPALEERKTDSYRPRGSKPKLALIAEFODSOG 328
Db 223 ---LAVAFONTVEPLIEFSYSDSIOYKRTAVVPRPKAGAVNPY--KFFVYNDLSLS 277
QY 329 KIVSTQEKELVOPSSLEPKVEYIARAGWTRDCKYAMAFELDRQOQLVLLPRLFTP 388
Db 278 SVTNATSIQIAPASMLIGD-HYLCDTWMTQERIS-----LQNL----- 316
QY 389 STENEDQRLASARAVPRVQYVY---YEEYVNWVINWHDIFPRPPOSEGDELCLFR 443
Db 317 -----RRIQNYSYMDICODDESSGRM----- 337
QY 444 ANECKTGRCHLYKTAVLKSQGYDMSPEPSCGE-----FK 480
Db 338 --NCLVARQHLKEMSTT-----GWGGRFPRSEPHFTLLDGNSEFKIISNEEGYRHICYFO 388
QY 481 CPIKEELATGEMEVILARHSKITVNEBETKLYFGCTKDPLEHNLVYVSYEAG--- 536
Db 389 IDKDCDFITGTMEVIGIEAL-----TSD-----YLXYSNEYKGMPGG 428
QY 537 -EIVRLTTPGFSHSCSMSONDFMVSHYSVSTPRCVHVKYK--SGP----- 560
Db 429 RNLVKIQLSDYTKVTKTSCLEINPERCQYYSVSFEKAQYQLRCSGGLPLTYLLHSSVD 488
QY 561 -----DDDLHKQPRWASMMEAASPRPYVPELFFHHTBSDVNL-IGMITPKPALQ 632
Db 489 KGLRYLEDNSALDK-----MLONVQ-----MPSKILDFILNETKFWYOMILPH-FD 535
QY 633 PGKKHPYLFYVGGPOVOLVANSFKIKYLTN---TLASIGYVUVY-IDGRSGCGRLR 688
Db 536 KSKRKLPLLDIYAGCSOKATV-----RLNATYTLASTENITIASFDRGSGYIGDK 569
QY 689 FEGALKNMGQVELEDDQEGE-QFVAEEYGEFIDSRAVIGHMSGCFGLSLMGLHKPQVE 747
Db 580 IMHAINRRLGTFEVEDQJEAROF--SKMGVDDKRIALMGMSYGGVYVTSWVLOSSGSYF 647
QY 748 KVALAGADVYMAVDTGTERYMDV--PENNOHGEAGVALHVKLPNEPNRLILHG 805
Db 648 KCGIYAVASWEYDYSYTERYMGLPREDENLDHRYSTWMSRAENF--KQVEYLLIHG 705
QY 806 FLDEVWHFHNFLVSQLIRGKRYQLODYIYNEHNSHTRCESGHHYEVLTLHFIOE 861
Db 706 TADDVWHFOQASQISKALVDYGVDFQAMMYTDEDHGLASSTAHQIYTHMSHTIKO 761

```

RESULT 4
US-08-619-280A-3
; Sequence 3, Application US/08619280A

[illegible]

```

Db 338 --NCLVANOHIEMSTT-----GWGRFRPSEPHFTLDGNSFKYIIISNEGRHICYFQ 388
QY 481 CPIKEEIALTSGEMEVLAHSGKIWNNEETKLVYFOGTKPTPLEHNLVVSYEAG---- 536
Db 389 IDKNCCTFITGTWTEVIGIEAL-----TSD-----YLXYSNEYKMGMG 428
QY 537 -EIVRLTTPGSHSCSMQNDMFVSHSYSTPCVHVYKL--SGP----- 580
Db 429 RNLKXIOISDVTYKVCISCELENPERCOYYSFSKKAQYQDLRCGSPGLPLTYLHSSVND 488
QY 581 -----DDDLHKQPRFWMASMEASCPDYPVPELFIHFTHRSQVRL-YGMITYKPHALQ 632
Db 489 KGLRVLEDNSALDK-----MLQNVQ-----MSKKIDFTILNETKFWYQMLLPH-FD 535
QY 633 PGKHPFVLVYGGPOVOLVNSSEFKIKYLRNLN--TLASLGAVVY-IDRSGSCQGLR 688
Db 536 KSKKYPILLDYAGPCSCAKDTVF-----RLNMTYLASTENITIVASFDRGSGYQGD 589
QY 689 FEGALKNQMGVELEDQVEGL-QVVAEKYGFIDLSRAIHSYSGGLSLMGLIHKPOVF 747
Db 590 IMHAINRRLTGFEVEDQIEAAROF--SKMGFVDNKRRIAMGMSYGGVYTSWVLGSGSGVF 647
QY 748 KVALAGAPVTWMAVDTGTYTERYMDV--PENNOHGEAGSVALHVEKLPNEPNRLILHG 805
Db 648 KCGIIVAVSHEWYDSYTERYMGLPTEPDNLHYRNSTVWSRAENF--KQVEYLILHG 705
QY 806 FLDENVHFHNLVLSOLIRAGKPYOLOIYNENHSHIRCPSEGEHYEVTLLHFOE 861
Db 706 TADNVHFOQSAQISKALVDYGVDFQAMWYTDHGHGIASTAHQIYTHMSHFITQ 761

```

RESULT 5

```

US-08-940-391-3
; Sequence 3, Application US/08940391
; Patent No. 5965373

```

GENERAL INFORMATION:

```

; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

```

COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: L0D 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; US-08-940-391-3

```

Query Match

```

Best Local Similarity 10.5%; Score 489.5; DB 2; Length 766;
Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;

```

```

QY 47 LINVKAPHPDFOVQGTDESGPHSHRLYYLGMPIYSGRENSLIYSELPKKRREALLLSWK 106
Db 26 VLKNG-----TDDATADSRKRYTL-----TDYKNTYRLKLYSLRW- 62
QY 107 QMLDHFQATPHHGVYSREELL-----REKRLGVFGITSTDPH-SSSGFLF 153
Db 63 -----ISDHELYLKQENNLIVFNAEYGNSSVPLENSTFDEFGHSINDYSISPDGQFIL 115
QY 154 QASNSLPHCRBGKNGEVPVSPMKPLEIKTQCSGRMPKICPAPDAF-FSFINSDMLVA 212
Db 116 LEVNVKQMRHSYTASTDYDILNKRQLITERIPNTQMTWSPVGHKLAVWNNDDLYK 175
QY 213 NIETGEERLTFCHQGLSNVLDDEKSAVATFVIOE-FDRFTGYWMCPTASWEGSEGLK 271
Db 176 IEPNLPYRIW--TGKEDIYN-----GITDYYEEEFSAVSLMWSNPTF----- 222
QY 272 TLRLIYEVEDESEVEVH---VPSPALEEKTDSEYRPRGSKNPKIALKAEFQDTSQ 328
Db 223 --LATAQFNDTEVPLEYSEFSDESLOQPKTVHVPYKAGAVNPTV--KFFVNTDLS 277
QY 329 KIVSTQEKELVOPSSLEPKVEYIARAGWTRDGKAYAMAFIDRQOMQLVLLPPLFIP 388
Db 278 SVTNATSIQITAPASMLGD-HYLCDVYATQERIS-----LQWL----- 316
QY 389 STENEQRLASARAVPRNVQYVY---YEENVNWINVDIYFPQSGEDELFLR 443
Db 317 -----RRIQNTSYMDICDYDESSGRW----- 337
QY 444 ANECKTGFCILYKVTAVIKSGDYMSSEFSGEDE-----FK 480
Db 338 --NCLVANOHIEMSTT-----GWGRFRPSEPHFTLDGNSFKYIIISNEGRHICYFQ 388
QY 481 CPIKEEIALTSGEMEVLAHSGKIWNNEETKLVYFOGTKPTPLEHNLVVSYEAG---- 536
Db 389 IDKNCCTFITGTWTEVIGIEAL-----TSD-----YLXYSNEYKMGMG 428
QY 537 -EIVRLTTPGSHSCSMQNDMFVSHSYSTPCVHVYKL--SGP----- 580
Db 429 RNLKXIOISDVTYKVCISCELENPERCOYYSFSKKAQYQDLRCGSPGLPLTYLHSSVND 488
QY 581 -----DDDLHKQPRFWMASMEASCPDYPVPELFIHFTHRSQVRL-YGMITYKPHALQ 632
Db 489 KGLRVLEDNSALDK-----MLQNVQ-----MSKKIDFTILNETKFWYQMLLPH-FD 535
QY 633 PGKHPFVLVYGGPOVOLVNSSEFKIKYLRNLN--TLASLGAVVY-IDRSGSCQGLR 688
Db 536 KSKKYPILLDYAGPCSCAKDTVF-----RLNMTYLASTENITIVASFDRGSGYQGD 589
QY 689 FEGALKNQMGVELEDQVEGL-QVVAEKYGFIDLSRAIHSYSGGLSLMGLIHKPOVF 747
Db 590 IMHAINRRLTGFEVEDQIEAAROF--SKMGFVDNKRRIAMGMSYGGVYTSWVLGSGSGVF 647
QY 748 KVALAGAPVTWMAVDTGTYTERYMDV--PENNOHGEAGSVALHVEKLPNEPNRLILHG 805
Db 648 KCGIIVAVSHEWYDSYTERYMGLPTEPDNLHYRNSTVWSRAENF--KQVEYLILHG 705
QY 806 FLDENVHFHNLVLSOLIRAGKPYOLOIYNENHSHIRCPSEGEHYEVTLLHFOE 861
Db 706 TADNVHFOQSAQISKALVDYGVDFQAMWYTDHGHGIASTAHQIYTHMSHFITQ 761

```

RESULT 6

```

US-09-794-236-1
; Sequence 1, Application US/09794236
; Patent No. 6337069

```



```

: GENERAL INFORMATION:
: APPLICANT: Groumann, Eric
: APPLICANT: Lacroix, Jean-Silvain
: APPLICANT: Monod, Michel
: TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
: FILE REFERENCE: 81985/276823
: CURRENT APPLICATION NUMBER: US/09/794.236
: CURRENT FILING DATE: 2001-02-28
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO: 1
: LENGTH: 766
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-794-236-1

Query Match          10.5%; Score 489.5; DB 4; Length 766;
Best Local Similarity 22.4%; Pred. No. 5e-40;
Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;

Qy 47 LTVNAPHDFOVOKTDESGPHSHRLYLGMFYSGRENSLXSELPKVKRKALLLSMK 106
Db 26 VLNKG-----TDDATADSRKTYTL-----TDYKNTYRLKLSLRW- 62
Qy 107 QMLDHFQATPHHGVYSREBELT-----REKRRLGVGITSYDFH-SESGLELF 153
Db 63 -----ISDHEYLKQENNILLVFNAKEYGNSSVPLENSTFDEFGHSINDYSISPGQFTL 115
Qy 154 QASNSLFHCROGKNGFVSPKPLEIKTQCGSPMDRPICPADPAF-FSFLNNSDLWYA 212
Db 116 LEYNVVKQMRHSYTAASYDIDLNKRQLTEERIPNNQWTVWSPGKHAAYVWNNDIYVK 175
Qy 213 NIETGEERLTCFCHOGLSVNLDDPKSAGATFVIOE-EDRFTGYWMCPTAMESEGLK 271
Db 176 IEPNLPSTRITW--TGKEDIYTN-----GITDWYEEVEVSAYSALMWSNGTF----- 222
Qy 272 TLRIIYEEVDESEVEYIH---VSPALERKTDYRYPRGSKNPALKAEFOTDSOG 328
Db 223 ---LVAQFNDEVEPLIEYSFSDESLOYPKTVRVYPKAGAVNPV- KFEVYNTDSLS 277
Qy 329 KIVSTQOEKELVOPFSSLEPKVEYIARAGWTRDGKYAMAFLDPRQOQLVLLPALFTP 388
Db 278 SVTNATSIQITAPASMLIGD-HYLCDVWATQERIS-----LCWL----- 316
Qy 369 STENEGORLASARAVPRNQPYVY-----YEEVTNWINVHDIIFYFPQSEGEDLCFIR 443
Db 317 -----RRIQNTSVMDICDYDESSGRW----- 337
Qy 444 ANECKTGFCHLYKVTAVLKSQGYDWSPPSPGDE-----PK 480
Db 338 ---NCLVAROHISMST-----GWVGRFRPSEPHFTLDGNSFYKLTISMEGCRHICYQ 388
Qy 481 CPIKEEIALTSGEWEVLARHGSKIWNNEETKLVPFOGTDPLEHHLYVVSYEAG- 536
Db 389 IDKDCOTFTKTGTWEYIGIEAL-----TSD-----YLYISNEYKMPG 428
Qy 537 -EIVALLTGTGFHSCMSQNFDMFVSHYSYSTPRCVNHYKL--SGP----- 580
Db 429 RNLYKIQLSDYTKVTCLSCLNPERCQYVSFSKRAKYQLRCSGDLPLTYLHSSVND 488
Qy 581 -----DDPLHKOPRFVASMMEASCPDPVPEIFHFHRSVRL-YGMVYKHAHQ 632
Db 489 KCLRVLEDSALDK-----MLQNVQ-----MPSKKLDFTILNEKFWYQMLPRH-FD 535
Qy 633 PGKHPFTVLYVGGPOVOLVNNSPKGIKYLRLN--TLASLGAAYV- IDRGSCQGRGLR 688
Db 536 KSKKPYLLIDYVAGPCSQADTYF-----RLNMATYLASTENIIYASPDGRSGYQGDK 589
Qy 689 FEGALKNONGUYEIEDQVAGL-QFYAEKGYFIDLSVALIHGNSYGFELSLMGLIHKPOYF 747
Db 590 IMHAINRRLGTREVEEDQIAARQF--SKMGFYDNKRRIALMGWSYGGYVSMVLGSSGVF 647
Qy 748 KVALIGAPVTVMMAVDYGTERTYMDV--PENNOHGYEAGSVALHVEKLPNEPRRLILHG 805

```

```

Db 648 KCGIYAVPSRWEYDYSYTERKMGIPTEPDLIDHRYNSTVMSRAENF--KQVEYLLIHG 705
Qy 806 FLDENVHFHFNFLVSQILRAGKPYQLOIYPRNERSINCPEGGEYEVTLHFLQ 861
Db 706 TADNVHFOQASQISKALVDGVDFQAMWYDDEHGIASTAHOHITYHMSHFIKO 761

RESULT 7
: Sequence 3, Application PC/TUS9307923
: GENERAL INFORMATION:
: APPLICANT: Morimoto, Chikao
: APPLICANT: Schlossman, Stuart F.
: APPLICANT: Tanaka, Toshiaki
: TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: COMPUTER: IBM PS/2 Model 502 or 55SX
: OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
: SOFTWARE: WordPerfect (Version 5.0)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/07923
: FILING DATE: 19930819
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/934,162
: FILING DATE: 21-AUG-1992
: APPLICATION NUMBER: 07/832,211
: FILING DATE: 06-FEB-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Frazer, Janis K.
: REGISTRATION NUMBER: 34,819
: REFERENCE/DOCKET NUMBER: 00530/055002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-5070
: TELEFAX: (617) 542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 755
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: PCT-US93-07923-3

Query Match          10.4%; Score 482.5; DB 5; Length 755;
Best Local Similarity 22.8%; Pred. No. 2.5e-39;
Matches 194; Conservative 116; Mismatches 321; Indels 219; Gaps 33;

Qy 93 KVKRKALLLSWKQMLDHFQATPHHGVYSREBELT-----REKRRLGVGITY 140
Db 39 KNTYRLKLSLRW-----ISDHEYLKQENNILLVFNAKEYGNSSVPLENSTFDEFGHS 90
Qy 141 SYDFH-SESGLELFQASNSLFFHCROGKNGFVSPKPLEIKTQCGSPMDRPICPADPA 199
Db 91 INDYSTSPGQFTLLEYNVVKQMRHSYTAASYDIDLNKRQLTEERIPNNQWTVWSPVG 150
Qy 200 F-FSFLNNSDLWYANITETGEERLTCFCHOGLSVNLDDPKSAGATFVIOE-FDRFTGYW 257
Db 151 HKIAYWNNDIYKIEPNLPSTRITW--TGKEDIYTN-----GITDWYEEVEVSAYSALM 204
Qy 258 WCPFTASWESGLKTLRIIYEEVDESEVEYIH---VSPALERKTDYRYPRGSKNPK 314
Db 205 WSPNGTF-----LVAQFNDEVEPLIEYSFSDESLOYPKTVRVYPKAGAVNPV 254

```


CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
FAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-230-491A-2

Query Match 9.2%; Score 426; DB 1; Length 760;
Best Local Similarity 23.6%; Pred. No. 1.3e-33;
Matches 168; Conservative 109; Mismatches 256; Indels 180; Gaps 31;

QY 195 PADPAFFSTINNSDLVANIETGEERRLTFCHQGLSNVLDLDPKSAQVATFVIOEEDRPT 254
DB 176 PGDPPE-----QITF--NGRENKIFN---GIPWYEEEM-LPT 208
QY 255 GY--WMCPTASWEGSEGLKTLRLIYEVDSEVEVIHVPSPALAE--RKTDSYRYPTGSK 311
DB 209 KYALMWSPPGKF-----LATAEFNDKDIPIVATSYGDEQYPRITINIPYPAGAK 258
QY 312 NPKIALKLAIEFOTDSOGKIVSTOEKELVOPFSSLPKV-----EYIARAGWTRDGKYAMA 366
DB 259 NPVARI-----FIIDTTYPAVYGPQEVPPAPAMIASDYFSSWLTWTVDKVC-- 305
QY 367 MELDRPOOWLQVLLRPALEFI-----PSTEN--EORL--ASARAARVWQY 410
DB 306 -----LQWLKRVQNVSVLSTCDFREDMOTWDCPKQEHIEESRTGAGGFVSRRVFSY 359
QY 411 VYEEVTNWIVNHIIFYPPOSEGEDELCEFLRANECKTGFCFLKVTAVLKSQGYDWS 470
DB 360 -----DAISYKIFSDKDGKRIHI----- 380
QY 471 PPSGDEDEKCPIKEETALTSGEWEVLARHSGKIWNNEETKLVYFOGT--KOTPLEHILY- 528
DB 381 -----KDTVENAIQITTSKWEAI-----NIFRYTODSLFSSNEEPEYGRNRIYR 426
QY 529 --VVSYEAGEIYRLTTPGFSHSCSM-----SONFDMFVSHSVSSTPPCVHYKLT- 577
DB 427 ISTGSIYPPSKCV-----TCHLRKRCQYTTASFSQYALVLCYGPPIPSTLH 477
QY 578 SGPDDEPLH--KOPRFWASMEASCPDPYVPEIHFHTRSDVRLKGMITKPHALQPGK 635
DB 478 DORTDEIKLEENKELENAKNIOLPKK-----EIKKLEVEDITLWYKMLPPO--FDRSK 532
QY 636 KHPYTLVYGGPOVOLVNNNSF--KGIKYLRNLTLASLGYAAVVYIDRGSCQORLREGAL 693
DB 533 KRPFLIQYGGCSQSVRSVFANWISYL-----ASKREGVIALVDRGAFQGDKILYAV 588
QY 694 KNOMGOVEIEDOEGLOFYAEKGYFIDLSRVALHGWSTGGLSLMGLIHKPOYFKYALAG 753
DB 589 YRKLGYEVEDOITAVRKFI--MGFIDERRIAITMGWSYGSYSSLASGTGFGKGIAY 647
QY 754 APYTVMAVDGTGTERYMNVP--ENNQHGVEAGSVLAHLEKLPNEBNRLILHGFIDENY 811
DB 648 APVSSWEYVYASVYTERFMGLPTKDDNLEHKNSTVMARAEYFRNV--YLLHGTADNV 705
QY 812 HFEHTFLYSOLIRACKPYOLOIYPPNERSIRCPESG---EHYEVTLHLEQ 861
DB 706 HFGNSQIAKALVNAQVDFQAMWYSDQNGH-----SGLSTNHLTYLHMTHTLKO 754

RESULT 10
US-08-619-280A-2

Sequence 2, Application US/08619280A

Patent No. 5767242

GENERAL INFORMATION:

APPLICANT: Zimmermann, Rainer; Park, John E.;

APPLICANT: Retlig, Wolfgang; Old, Lloyd J.

TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN

TITLE OF INVENTION: ALPHA, AND USES THEREOF

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESS: Felfe & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/619,280A

FILING DATE: 18-MARCH-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/230,491

FILING DATE: 20-APRIL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 5767242man D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5330.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 760 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-619-280A-2

Query Match 9.2%; Score 426; DB 1; Length 760;
Best Local Similarity 23.6%; Pred. No. 1.3e-33;

Matches 168; Conservative 109; Mismatches 256; Indels 180; Gaps 31;

QY 195 PADPAFFSTINNSDLVANIETGEERRLTFCHQGLSNVLDLDPKSAQVATFVIOEEDRPT 254
DB 176 PGDPPE-----QITF--NGRENKIFN---GIPWYEEEM-LPT 208
QY 255 GY--WMCPTASWEGSEGLKTLRLIYEVDSEVEVIHVPSPALAE--RKTDSYRYPTGSK 311
DB 209 KYALMWSPPGKF-----LATAEFNDKDIPIVATSYGDEQYPRITINIPYPAGAK 258
QY 312 NPKIALKLAIEFOTDSOGKIVSTOEKELVOPFSSLPKV-----EYIARAGWTRDGKYAMA 366
DB 259 NPVARI-----FIIDTTYPAVYGPQEVPPAPAMIASDYFSSWLTWTVDKVC-- 305
QY 367 MELDRPOOWLQVLLRPALEFI-----PSTEN--EORL--ASARAARVWQY 410
DB 306 -----LQWLKRVQNVSVLSTCDFREDMOTWDCPKQEHIEESRTGAGGFVSRRVFSY 359
QY 411 VYEEVTNWIVNHIIFYPPOSEGEDELCEFLRANECKTGFCFLKVTAVLKSQGYDWS 470
DB 360 -----DAISYKIFSDKDGKRIHI----- 380
QY 471 PPSGDEDEKCPIKEETALTSGEWEVLARHSGKIWNNEETKLVYFOGT--KOTPLEHILY- 528
DB 381 -----KDTVENAIQITTSKWEAI-----NIFRYTODSLFSSNEEPEYGRNRIYR 426
QY 529 --VVSYEAGEIYRLTTPGFSHSCSM-----SONFDMFVSHSVSSTPPCVHYKLT- 577
DB 427 ISTGSIYPPSKCV-----TCHLRKRCQYTTASFSQYALVLCYGPPIPSTLH 477

```

OY 578 SGRDDPPRH--KORFPMASMEASACPDYVRELFPHHTNSDVRALGMUYKRALOPK 6353
Db 478 DGRDQDEKILEENKELENALKNIOLPRE----ELKLENDETITMYKMLPPD-FDRSK 5322
OY 636 KHPYLVENYGRPOVOLYNNSF--KGIKYLRINTLASLGAYAVVYIDGRSGOSGRLEFGAL 6939
Db 533 KYPLLIQYGGPCSOVSYSVAVMNISTL---ASKGAMVATAYDGRGTFQOGRKLLYAV 5888
OY 694 KNOHQVIEIEOVESLOEVAKYEFGIDLSRVALHSGWYGGGLSLMGLHKPOYKVALIAG 7533
Db 589 YRKLGUYVEEOITRAVRFTE-MGFIDBKRAIAGWSTIGGVSSLLASLSTGTEFGCIIAV 6477
OY 754 APYTVMAAYDTGYTERYMDV--ENNOHGVEAGSVALHVEKLPNEPRRLILHSGLENDY 8111
Db 648 APVSMEWYASVYTEREFGTLPTKDDNLEHYNKSTPYMARAEYFRNVD--YLLHGTADNV 7055
OY 812 HEPHTNPLVSOILRAGKRYOQIYPRNHRHSRCESG---EHYEVTLIHLHLOE 861
Db 706 HFQNSAQIAKLALVNAOYDFQAMWYSDNNHGJ---SGLSITNHLIYTHNTHELKKO 754

```

```

Db 109 YGVSKPL-----WSP-----DGSIIYITISLGESESIDR-- 138
QY 286 EVIHPSPALERKTDYRYPRTGSKNPKIALKLAEPOTDGG-----KIYST 333
Db 139 -----KKTEDDSYE-----PVEQGLSYKRDGGLTRGAYADLVVSVKSG 179
QY 334 QEKELVOPFSSLFPVEYIARAGWTRDGKYAMAMFLDRPQOMQLVLLPPLALFISTENE 393
Db 180 EMKELTSS-----HKADH-GDPAFSPDGK--WLVF-----SANTLT 210
QY 394 EORLASARAVPRNVOPIVVEEVTNWINVDIFYPPQSGEDELCEFLRANECKTGFC 453
Db 211 ETDAS-----KPHDVY-----IMLSGDLK----- 232
QY 454 LYKTAVALKSGCYDMSEFSP-----GEDEKCPKEKRIALTSGEM----- 494
Db 233 -----QVTPHKGSGSSSPDGRYLLALGNKEKYK-----NATLSKAMLYDIEQRLT 281
QY 495 ---EVLARH-----GSK---IWNVEETKLVEFGOTKDPPLREHILYVSYEAG 536
Db 282 CLTEMLDVHLADALIGDLIGAGBQRPIM-TKDSGIFYVIGDQG--STGIYIIEGLV 338
QY 537 EIVRLTFPGSHSCSMQNFDMFVSHSYSTPCVHVYKILSGPDDPLHKQPRFMA5MM 596
Db 339 YPIRLERK-YINSFSLSPDEQHFIA5YTKPDRPS--ELYSI-----PLQGEKQLTGAN 389
QY 597 EAASCPPIVPEIFHFTRSDYRLXGMITKPHALQPKKHPTVLFVGGPVOVLVNSF 656
Db 390 DKFVREHTISIPETIOYATEDGVNVMGLMRPOMEGETTYPLILINHGPGIMMGHYTF 449
QY 657 KGIKYRLNTLASLGYAVVVIDGRSGCGRLFEKALKNQOMQVELEDOVELOFAEKY 716
Db 450 H-----EFOVLAAKYIAVYINPRSGHSGQEFVNAVRGDYDGKRYDVMQAVDAIKRD 504
QY 717 GTIDLSRAVHIGW5YGGFLS-LMGLHKPOYKVALIAGAPYVWMAV---DTGY--TE 768
Db 505 PHIDKRLRGVGTSGYSGFMTNMIYQOTNR---FKAAYTORISNMISFYGVDIGYFED 561
QY 769 RYM--DVEENNQHGVEASV--ALHVEKLPNPNRLLILHGLDENVHFHFNFLVSQI 824
Db 562 WOLEHDMEDFETKLMRSPKYAANVE-----TPLLILHGERDRCPLEQAEQLFALK 615
QY 825 RAGKPYQLOIYPNERHSI 842
Db 616 KMGKETKLVPRFNASHNL 633

```

```

RESULT 13
US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshiyugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACTL PEPTIDE HYDROLASE AND GENE ENCODING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; EARLIER FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

```

```

Query Match 5.28; Score 240.5; DB 4; Length 632;
Best Local Similarity 20.8%; Pred. No. 4.9e-15;
Matches 158; Conservative 105; Mismatches 272; Indels 225; Gaps 36;

```

```

QY 132 KRLGVFGITTSDFHSE---SGFLFQASNSLFLHCHDGGKNGFWSPMKPLEIKQCSOPR 188
Db 23 KGAIVQVTEISLKDDDFSKLYLY-----DGKR-----VKPFTSGMNSMR 65
QY 189 MDPKICPADPAFESFIN-----NSDLVANIENGGEERRLTFCHQGSNV--LDDPKSA 239
Db 66 FSP-----NGKLIATFSKRDKGEKSELYVLPDGGKRLLAKEKYGIKILRFTEDGKSI 120
QY 240 GVAT-----FVIOEFDRFTGYWMCPTASWEGSLKTLRIIYEEVDESEVEV 287
Db 121 AVVPIPIDEKKGNDVHLIRLPEFNGVGW-----IYGR--RNMYVL 161
QY 288 IHVPSPALERKTDYRYPRTGSKNPKIALKLAEPOTDGGKIYST--QEKELVOPFSSL 345
Db 162 VDV-----ESGKKRRLPKN---LNVDOI RFHN--GRLYFTAOEDRERKPLISDL 206
QY 346 F---PKVEYIARAGWTRDGKYAMAMFLDRPQOMQLVLLP--PALFIPSTENEBORLAS 399
Db 207 YLENRRKVRKLT-----DEKMIPLFLPLDGSFVLKANLTLE----- 243
QY 400 ARAVRNVOPIVVEEVTNWINVDIFYPPQSGEDELCEFLRANECKTGFCILYKTA 459
Db 244 -RGITPNH-----IYHDP-----KTG--ELKKLTK 267
QY 460 VLKSGCYWSEFSPGEDEKCPKEKRIALTSGEMVLAHSGKTIWNEETKLVEFG-- 517
Db 268 DLDRNAVN-----SLNSDVRSQRAELVYKEGMI 296
QY 518 ---TKDTPLEHLLVYVSYEAGEIYRLTPPGSHSCSMQNDMEVSHSYSTPCVHY 574
Db 297 YVATDGP--RAVLFVYVNDL--GKIERVI--GGRDSVESEFDIDYLAFTAQDAVTELETL 351
QY 575 YKLSGPDPLHKQPRF--WASMEASCPDPYVPEIFHFTRSDYRLXGMITKPHALQ 632
Db 352 YR-----DGEKKKYVDPMKMIKGYTLK-----PEHKVKAASGVLEIDAVMMPVNER 399
QY 633 PEKHPYLVFYVGGQVOVLVNSFQIKYL--RLNTLASLGYAVVVIDGRSGQRLRREG 691
Db 400 KKKKYPALLEIHGPKRTAY-----GYAFMEHFHVLTSKGFVYIINPRSGSDGYDEEP--A 452
QY 692 ALKNQMGVEITDOYEGLOFAEKYGTIDLSRAVHIGW5YGGFLS--LMGLHKPOYKRV 749
Db 453 DIRGHYGERDYDLMVEYDEALRRDFIDGERLGVGTSGYSGFMTNMIYGHTR---FKA 509
QY 750 A1AGAPYVWMAV---DTGY---TERYMDVPEENNQHGVEASVALHVEKLPNPNRLLI 802
Db 510 AVTQRSISMWISFFGTIDIGYFAPDQIGKDPMSNLEGYWEKS---PLKYAVNVEFPLLI 566
QY 803 LHGFLDENVHFHFNFLVSQIRAGKPYQLOIYPNERHSI 842
Db 567 IHSTEDYRCMLPEALQFLISLTKYIGKRVLAIFPENHDL 606

```

```

RESULT 14
PCT-US93-07923-11
; Sequence 11, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 593
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
PCT-US93-07923-11

```

```

Query Match      4.58; Score 207; DB 5; Length 593;
Best Local Similarity 18.9%; Pred. No. 1.1e-11;
Matches 136; Conservative 93; Mismatches 255; Indels 234; Gaps 31;

```

```

47 LIVKAPHDEQFVQKTESGPHSRILYLGMPYSGRENSLLYSEIPKVKREALLLSMK 106
26 VLLKKG-----TDDATADSKRYTYL-----TYLKNTYKLYKLSLKW- 62
107 QMLDFOATPHHGVYSREELL-----REKRLGVFGITSYDFH-SESGLEPLF 153
63 -----ISOHELYLKQENNLIVFNAEYGNSSVLENTPEFGHSINDYISPDGQITL 115
154 QASLSLPHCRDGGKNGEYVSPMKPLEIKTQCSGPRMDPKICPADPAF-FSFINNSDLWA 212
116 LEYVYVQWHRSHSYASYDIYDLNKRROLITEERIPNNTQWYVSPVGHKLAVYMNNDIYVK 175
213 NIEGERRRLFFHOGISLVNDDPKSAGVATFVIOEE-FDFFTGYVMCPTSMGSEBGL 271
176 IEPNLPSTRIWT--TGKEDIYN-----GIDWVYEEVFSAVSLMWSPNCTF----- 222
272 TLRLIYEVEDESEVEVH-----VPSPALERKTDGYRPRTSKNPKIALKLAEFOTDSOG 328
223 ---LAVQFNTEVPLIEFSYSDESIQPKTVAPYPRKAGAVNPTV--KFFVYNTDSLS 277
329 KIYSTOKELVOPSSLFPRKVEYIARAGWTRDGKYAMAMFLDRPOQWLQVLLPPLAETP 388
278 SVTWTATSIQITAPASMLIGD-HYLCDWYATQERIS-----LQWL----- 316
389 STENEGRRLASARAVPRNVPYV-----VEEYVNWVIMVNDIYFPFQSGEDELCLRL 443
317 -----RRIONYSVMDICDYDESSGRW----- 337
444 ANECKTGFCHLYKTVAVLSQGYDMSEPFSGEDE-----FK 480
338 --NCLVARQHHEKMT-----GWGGRFRSEPHFTLDGSPFKIITSNENRGYHICFYQ 388
481 CPIKEEIALTSGEWEVLARHGSKTIWNEETKLYVFQGTKDTPLDHLHYVVSYEAG---- 536
389 IDKDCFTFIKGTWEVIGIEAL-----TSD-----YLYYISNEXKMGHGG 428
537 -EIVRLTTPGFSHSCSNQNDMFVSHYSYSPPCVHYVYL--SGP----- 580
429 RNLYKIQLSDTYKVTCLSCELNPERCOYYSVSFSKAKYQOLRCSGGLPLYLTHSSVND 488

```

```

581 -----DDDLPHKOPRFWASMEASCPDPYVPEIFHFHTRSDVRL-YGMIXKPHALQ 632
489 KGLRVLEDNSALDK-----MLQNVQ-----MPSKADFIILNETKRWYGMILRPH-FD 535
633 PGKHPVLYEVYGGPOVOLVNNSEKGIKYLRLN--TLASIGYAVVY-IDGRGSCQRC 686
536 KSKRYPLLDVYAGPSCQKADTVF-----RLNMATYLASTENIIVASFDRGSGYQG 587

```

RESULT 15

```

US-09-390-234-20
Sequence 20, Application US/09390234
Patent No. 6365390
GENERAL INFORMATION:
APPLICANT: Blum, David L.
APPLICANT: Katsava, Irina
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Phenolic Acid Esterases, Coding Sequences and Methods
FILE REFERENCE: 67-98
CURRENT APPLICATION NUMBER: US/09/390,234
CURRENT FILING DATE: 1999-09-03
EARLIER APPLICATION NUMBER: US 60/099,136
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 721
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-09-390-234-20

```

```

Query Match      3.5%; Score 163.5; DB 4; Length 721;
Best Local Similarity 21.5%; Pred. No. 3.7e-07;
Matches 98; Conservative 56; Mismatches 193; Indels 107; Gaps 20;

```

```

467 DWSEP-FSPGDE-----FKCPIKEEIALTSGEWEVLARHGSKIW 505
289 DSSSPVSPNDKLAIPQMRDEYESDRALLVYVSLGSKKTIIPVAGDWD--RSPDGV- 344
506 VNEETKLYVFOGTYDTPLEHLLVYVSYEAGEIVRLTTPGFSHSCSNQNDMFVSHYSS 565
345 -----KWTLP-DGKTLIVGSEDLGRTRLSIPANAKDDYKPKNF-----TDGGS 386
566 VSPPCVHYVYKLSGRDDPLHKOPRFWASMEASCPDPYV-----P 607
387 VSA-----YFLL--PDSSLVTGSALMTNMVYTAAPKPKGVYIKKIASANEIDPELKLGP 439
608 PEIFHFTRSD-VRLGYMIYKPHALQPGKHPVLFVYGGPOVOLVNNSEFGIKYLRNL- 665
440 SDISEFYQGNFTDIHNAVITPENFDKSKTYPLFLFIHGQ-----GWMADGNS-TRNRP 494
666 -TLASIGYAVVYIDGRGSCQRCGLRFEGLAKNOMGOVEIEDOVESGLQVAEKYGIDLSRV 724
495 KAMADOGVVVAPRPFTSGFGQALTTAIONMGCAPYDDLVKCEWYVHENIADVDDHG 554
725 AIHMGSYGGL-----SLMGLHKRPOYK--VAIAGAPV--YMAAYDGYTERMDV 773
555 VAAAGSTGEGMINWIOGSPGAKRKALVSHDGTVDADKYSTBELWP-MOREFGFTWDA 613
774 PENNOHGEAGSVALAHEKLEPNEPNR-----LLILHGLDENVHFHTFLVSLQIRA 826
614 RDN-----YRRND-----PSAPERILQFTPTPLVLIHSDQDYDLPLVAEGLSLFNVLQER 661
827 GKPYOLOIYVNERHSIRCPESGEHYEVTLLHFOEY 862
662 GVPSRFLNFPDENHNVVNPENSLVWMOALGWINKY 697

```

Search completed: December 12, 2002, 12:07:17
Job time : 24 secs.

DR Pfam: PF00326; Peptidase_S9; 1.
DR PRINTS: PR00793; PROAMOPTASE.
SQ SEQUENCE 863 AA: 98263 MW: 40FE0B7BE26CDED5 CRC64:

Alignment Scores:

Pred. No.:	1, 83e-301	Length:	863
Score:	4646.00	Matches:	863
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.89%	Indels:	0
DB:	4	Gaps:	0

US-09-976-674-4 (1-2617) x Q8WXB8 (1-863)

```
QY 11 ATGGCCACACCGGAGCCCAACAGCCGAGCGAGCGAGCGGACGATGACCG 70
   |||
Db 1 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspAspPro 20

QY 71 GCGGCGCGCTTCAGAGTGCAGAAACACTCGTGGAGCGGCTCCGAGCATCATCCAGGC 130
   |||
Db 21 AlaAlaArgPheGlnValGlnLysHisSerThrPaspArgLeuArgSerIleIleHisGly 40

QY 131 AGCGGCAAGACTCGGCGCTCATTTGTCAACAAGCGCGCCGACGACTTCAGTTGTGCAG 190
   |||
Db 41 SerArgLysThrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60

QY 191 AAGAGGATAGTCTGGCGCCCACTCCACCGGCTTACTACCTGGGAATGCCATATGGC 250
   |||
Db 61 LysThrAspLysSerGlyProHisSerHisArgLeuThrLysLeuGlyMetProTyrGly 80

QY 251 AGCGGAGAACTCCCTCTCTACTGTAGATTCCCAAGAAGTCCGGAAGAAGGCTCTG 310
   |||
Db 81 SerArgLysAsnSerLeuLeuTyrSerGluIleProLysValArgLysGluAlaLeu 100

QY 311 CTGCTCTGTCTCGGAAGCAGATGCTGGATCATTTCCAGGCCAGCGCCACATGGGCTC 370
   |||
Db 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120

QY 371 TACTCTGGAGAGAGAGTCTGTAGAGGAGCGGAACGCTGGGGGCTTCGCGCATCAC 430
   |||
Db 121 TyrSerArgGlnGlnGlnLeuLeuArgGlnArgLysArgLeuGlyValAlaPheGlyIleThr 140

QY 431 TCTTACGACTTCCACAGCGAGAGTGGCTCTCTCTTCAGGCGCAGCAAGGCTTTC 490
   |||
Db 141 SerTyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160

QY 491 CACTGTGCGCAGCGCGCCAGAAAGCGCTTCATGCTGCCCTTAAGAAACCGCTGGAATC 550
   |||
Db 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 180

QY 551 AAGACCCAGTGTCAAGGCGCCGAGTGGAGCCCAAAATTCGCCCTGCGGACCGCTTC 610
   |||
Db 181 LysThrGlnLysSerGlyProArgMetAspProLysIleLysProAlaAspProAlaPhe 200

QY 611 TTTCTCTTCAATTAACAGCGACTGTGGTGGCCAAATCGAGACAGCGAGGAGCGG 670
   |||
Db 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGlnThrGlyGlnLysArg 220

QY 671 CGGCTAGACCTTGTGCCCAAGGTTTATCCAAATCTCTGGATGACCCCAAGTCTGGCGGT 730
   |||
Db 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240

QY 731 GTGGCCACCTTGTGCATACAGAAAGATTGCAGCGCTTCAGTGGGTACTGTGGTGGCCG 790
   |||
Db 241 ValAlaThrPheValIleGlnGlnGlnPheAspArgPheThrGlyTyrTrpTrpCysPro 260

QY 791 ACAGGCTCTGGGAAGGTTCAAGAGGCTCAAGAGCGCTGGGAATCTGTATGAGGAATC 850
   |||
Db 261 ThrAlaSerTrpGlnLysSerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnGlnVal 280

QY 851 GATGAGTCCGAGGTGGAGGATTCACAGTCCCTCTCTGCGGTGAGAAAGAAAGAAACG 910
   |||
Db 281 AspLysSerGlnValGlnValIleHisValProSerProAlaLeuGlnGlnArgLysThr 300
```

```
QY 911 GACTCGTATCGGTACCCCGAGACAGCGCAAGAAATCCCAAGATTGCGCTTGAACCTGGCT 970
   |||
Db 301 AspSerTyrArgTyrTrpAlaGlyThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320

QY 971 GACTTCAGACTACACAGCGGCAAGATCGTCTGCAGCCAGAGAAAGAGTGTGCAG 1030
   |||
Db 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnLysGlnLeuValGln 340

QY 1031 CCTTACGCTCGCTGTTCGCCGAAGGTGAGTAAATCGCCAGCGCGCGGTGGACCCGGGAT 1090
   |||
Db 341 ProPheSerSerLeuPheProLysValGlnTyrIleAlaArgAlaGlyTrpThrArgAsp 360

QY 1091 GGCAAATACGCTTGAGCCATGTCCTGAGCGGCGCCAGCAGATGAGTCCAGTCTGCTC 1150
   |||
Db 361 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnIleThrPheGlnLeuValLeu 380

QY 1151 CTCCCGCGGCGCTGTTCATCCGAGCAGCAAGAAATGAGCAGCGGCTAGCCCTGCGC 1210
   |||
Db 381 LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGlnArgLeuAlaSerAla 400

QY 1211 AGAGCTGTCCCGAAGATGTCCAGCGCTATGTGTGTACGAGAGGTCCACCAAGTCTGG 1270
   |||
Db 401 ArgAlaValProArgAsnValGlnProTyrValValTyrGlnGlnValThrAsnValTrp 420

QY 1271 ATCAATGTTCATGACATCTTATCCCTTCCCAATCAGAGGAGAGAGAGTCTGTC 1330
   |||
Db 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGlnGlyLysAspGlnLeuCys 440

QY 1331 TTTCTCGCGCCCAATGAATGATCAAGACCGGCTTCCCATTTGTACAAAGTACCGCGCTT 1390
   |||
Db 441 PheLeuArgAlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460

QY 1391 TTTAAATCCCAAGGCTTCAGTTGAGTGCAGCCCTTCAGCCCGCGGGAATGATTAAG 1450
   |||
Db 461 LeuLysSerGlnGlyTyrAspTrpSerGlnProPheSerProGlnGlyLysAspGlnPheLys 480

QY 1451 TGCCCACTTAAGAAAGATGTGCTTACACAGCGGTAAGGAGAGTGTGGCGAGGCAC 1510
   |||
Db 481 CysProIleLysGlnGlnIleAlaLeuThrSerGlyLysTrpGlnValLeuAlaArgHis 500

QY 1511 GGTCCAAAGATGTGGGTCAATGAGAGACCAAGCTGTGACTTCCAGGCGCCCAAGGAC 1570
   |||
Db 501 GlySerLysIleTrpValAsnGlnGlnThrLysLeuValTyrPheGlnGlnTyrThrLysAsp 520

QY 1571 AGCGCGGTGAGACCAACCTTACGTGTACGATGAGCGGCGCGCGAGATCGTACGC 1630
   |||
Db 521 ThrProLeuGlnHisHisLeuTyrValValSerTyrGlnAlaIleGlyIleValArg 540

QY 1631 CTCACACGCGCGGCTTCTCCATAGCTGTCCATGAGCCAGCAACTTGCATGTTCTGTC 1690
   |||
Db 541 LeuThrThrProGlnPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560

QY 1691 AGCCACTACAGCAAGCTGAGCAGCGCGCTGCTGCTGCAAGCTTCAAGAGTGAAGGCGCCC 1750
   |||
Db 561 SerHisTyrSerSerValSerThrProCysValHisValTyrLysLeuSerGlyPro 580

QY 1751 GAGGAGACCGCGTGCACAGAGCGCGCTCTGCGGCTGAGTGAAGGAGGAGCGACG 1810
   |||
Db 581 AspAspAspProLeuHisLysGlnProArgPheThrAlaSerMetCysGlnAlaAlaSer 600

QY 1811 TGCCCGCGGATTAATGTTCTTCCAGAGATTTCCATTTCCACAGCGGCTTGGATGTCCGG 1870
   |||
Db 601 CysProProAspArgValAlaProProGlnIlePheHisAspHisThrArgSerAspValArg 620

QY 1871 CTCTACGGCATGATTAACAAGCCCAAGCGCTTGCAGCGGAGAAAGAACCCACCGCTC 1930
   |||
Db 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlnLysLysHisProThrVal 640

QY 1931 CTCTTGTATATGAGGCGCGCGAGTGCACAGTGTGAATGAACTTCAAGGACATCAAG 1990
   |||
Db 641 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
```



```

QY 1991 TACTTGGCGCTCAACACACATGCGCTCCCTGGGCTACGCCGTGTGATTGACGGCAG 2050
    |||||||
Db 661 TyrlleuArglleuAsnThrleuAlaSerleuGlyTrlAlaValIleAspLylArg 680
QY 2051 GCGTCTCTCAGCGAGGCTTCGGTTCAGAGGGCCCTGAAACCAATGGCCAGGTG 2110
    |||||||
Db 681 GlySerCysGlnArgLylleuArpHegLuglAlaLeuLysAsnGlnMetLylLlnAl 700
QY 2111 GAGATCGAGGACCGAGGAGGCGCTGAGTTCCTGGCCGGAACATATGCTTCAATCAG 2170
    |||||||
Db 701 GlutIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTrglYpHelleasp 720
QY 2171 CTGAGCCGAGTGGCCATGCTGCTGCTACAGGGGCTTCCTGCTATGAGGCTA 2230
    |||||||
Db 721 LeuSerArgValAlaIleIleHisGlyTrpSerTrglYglYpHelleuSerleuMetLylLeu 740
QY 2231 ATCCACAAAGCCCGAGGTTCAGAGTGCATCGCGGGTGCCTCCGCTGATGATG 2290
    |||||||
Db 741 IleHisLysProGlnValAlaPheLysValAlaIleAlaGlyAlaProValThrValTrpMet 760
QY 2291 GCGTACGACAGAGGTACAGTACGAGCGGTACATGAGCGTCCCTGAGAAACACAGCAGCGC 2350
    |||||||
Db 761 AlaTrpAspTrlHglYTrlThrGluArgTrlMetAspValProGluAsnGlnHisGly 780
QY 2351 TATGAGCGGGTCCGCTGCGCTGACAGTGGAGAGTGCCTGATGAGCCCAATGAGCCCAACGCTTG 2410
    |||||||
Db 781 TyrGluAlaGlySerValAlaIleHisValGluLysLeuProAsnGluProAsnArgLeu 800
QY 2411 CTATATCTCCACAGGCTCTCTGAGCAAAAGTGCATCTTTTCCACACAACTCTCTGTC 2470
    |||||||
Db 801 LeuIleLeuHisGlyLysPheLeuAspGlnAsnValHisPheHisThrAsnPheLeuVal 820
QY 2471 TCCCAACGATACCCGAGCGAGGAAACCTTACACGCTCCAGATATACCCCAACAGAGACAC 2530
    |||||||
Db 821 SerGlnLeuIleArgAlaGlyLysProTrglLeuGlnIleTrpProAsnGlnArgHis 840
QY 2531 AGTATTCCTGCGCGAGTGGCGGAGCAGCATATGAGTCACGCTTCTGATCTTCTACAG 2590
    |||||||
Db 841 SerIleArgCysProGlnSerGlyLysIleHisTrglValThrLeuLeuHisPheLeuGln 860
QY 2591 GAATACCTC 2599
    |||||||
Db 861 GluTrpLeu 863

RESULT 2
QYDAG6 PRELIMINARY; PRG: 883 AA.
ID Q9DAG6
AC Q9DAG6;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2002 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE 4932434F09RIK.
GN 4932434F09RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090.
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Sato T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Botfeill D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

```

```

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weiss C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK016546; BAB30295.1; -.
DR MGD: MGI:1921638; 4932434F09RIK.
DR InterPro: IPR002469; DPPIV_N_term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; Ser_estr_site.
DR Pfam: PF00930; DPPIV_N_term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
SQ SEQUENCE 883 AA; 101297 MW; EA958D6BD33EC71F CRC64;

Alignment Scores:
Pred. No.: 3,83e-184 Length: 883
Score: 2887.50 Matches: 520
Percent Similarity: 76.72% Conservatve: 139
Best Local Similarity: 60.54% Mismatches: 197
Query Match: 60.22% Indels: 3
DB: 11 Gaps: 3

US-09-976-674-4 (1-2617) x Q9DAG6 (1-883)
QY 26 ACCCCACGCGCCGAGCGAGCGAGCGAGCGCCGACAGATGACCGCGCGCCG---TTC 82
    ||| :||| :||| :|||
Db 17 ThrAlaGluCysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 36
QY 83 CAGGTGCAGAAACACTGCTGGGAGCGGCTCCGAGACATCATCCAGCGAGCGCGAAGTAC 142
    ||| :||| :||| :|||
Db 37 TyrValAlaGluGlySerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyr 56
QY 143 TCGGGCCATATGTCAACAAAGCGGCCACGACGCTTCCAGTTGTGCAGAAAGGAGATGAG 202
    ||| :||| :||| :|||
Db 57 HisGlyTrpMetMetAlaLysAlaProHisAspPheMetPheValLysArgThrAspPro 76
QY 203 TCTGGGCCCCACTCCACCGCCTCTACTACTAGATGAGATGCCATATGAGCAGCGGAGAAC 262
    ||| :||| :||| :|||
Db 77 AspArgProHisSerAspArgValTrpTyrLeuAlaIleMetSerGlyGluAsnArgGluAsn 96
QY 263 TCCCTCTCTACTGTGAGATTCCCAAGAAAGTCCGGAAGAGGCTCTGCTCTCTCTCC 322
    ||| :||| :||| :|||
Db 97 ThrLeuPheTyrSerGlnLeuProLysThrIleAsnArgAlaAlaValLeuMetLeuSer 116
QY 323 TGGAGCAGATGCTGATCATTTCCAGGCGCAGCCCAACCATGGGGGTCTACTCTGGGAG 382
    ||| :||| :||| :|||
Db 117 TrpLysProLeuLeuAspLeuPheGlnAlaTrpLeuAspTrglYMetTrpSerArgLys 136
QY 383 GAGGAGCTGCTGAGGAGCGGAGAAAGCCCTGGGGTCTTGGCATCACTCTACGACTTTC 442
    ||| :||| :||| :|||
Db 137 GluGluLeuLeuArgGluArgLysArgIleGlyThrValGlyIleAlaIleArgLysPyr 156
QY 443 CACACGAGAGTGGCTCTCTCTCTTCCAGGCGCAGCAACAGCCTCTTCCACTGTGGCAG 502
    ||| :||| :||| :|||
Db 157 HisProGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrIleLysLys 176
QY 503 GCGCGCAAGAACGGCTTTCAGTGGTGTCCCTATGAACCGCTGGAATCAAGACCCAGTGC 562
    ||| :||| :||| :|||
Db 177 GlyGlyProHisGlyPheThrGlnGlnProLeuAlaGProAsnLeuValGluThrSerCys 196
QY 563 TCAGGCGCCCGAGATGAGCCCAAAATCTGCTCCCTGCGGACCGCTCTTCTCTCTTCATC 622
    ||| :||| :||| :|||
Db 197 ProAsnIleArgMetAspProLysLeuCysProAlaAspProAspTrpIleAlaPheIle 216
QY 623 AATTAACGAGGACTGTGGTGGCCAAACATCGAGACAGGAGGAGGAGCGGCGCTGACCTTC 682
    ||| :||| :||| :|||
Db 217 HisSerAsnAspIleTrpIleSerAsnLeuValThrArgGluGluArgArgIleThrTyr 236
QY 683 TCCCAACGAGGTTTATCCAAATGCTCTGATGACCCCAAGTCTGGGGTGTGGCCACCTTTC 742

```

```

Db      237 ValHisnngluLeuAlaAsmGluGluAspProAlrGSerAlaGluValAlaThrPhe 256
OY      743 GTCTACAGGAAGAGTTGCGAGCTTCAGTGGTACTGGTGGTCCCGCCAGACGCTCCGG 802
Db      257 ValLeuGlnGluGluPheAspArgTyrSerLysTyrTrpTrpCysProGlnAlaGluArg 276
OY      803 GAAGGTTCAAGAGGCGCTCAAGACGCTCGAATCCCTATGAGAAAGTGCATGAGTCCGAG 862
Db      277 ThrProSerGlyLys--LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGlu 295
OY      863 GTGAGGTCATTCACGTCCTCCCTCCGCTGAGACAAAGGAAGAGCGAGCTATCCG 922
Db      296 ValGluIleIleHisValThrSerProMetLeuGluThrArgArgAlaAspSerPheArg 315
OY      923 TACCCAGACGAGCGACAGAAATCCCAAGATGCTTGAATCGGGGAGTTCACAGCT 982
Db      316 TyrProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleValVal 335
OY      983 GACAGCCAGGCAAGATCGTCTCGACCCAGAGAAAGAGAGCTGTGACGCCCTTCAGCTCG 1042
Db      336 AspAlaAlaGlyIleIleAspValIleAspLysGluLeuValGlnProPheGluIle 355
OY      1043 CTGTTCCCGAAGGTGAGTACATGCCAGAGCGCGGTGGACCCGGATGCGAAATACGCC 1102
Db      356 LeuPheGluGluValGluTyrIleAlaArgAlaGluTyrTrpThrProGluGluLysHisAla 375
OY      1103 TGGGCGCATGCTCGAGACGGCGCCCGGACAGTGGCTCCAGGCTGCTCCCTCCCGCGGCC 1162
Db      376 TrpSerIleLeuLeuAspArgSerGlnThrHisLeuGlnIleValLeuIleSerProGlu 395
OY      1163 CTGTTCATCCCGAGACAGAAATAGAGAGAGAGCGAGCTAGCGCTTCGACAGAGCTGCC 1222
Db      396 LeuPheIleProValGluAspAspAlaMetAspArgGlnArgLeuIleGluSerValPro 415
OY      1223 AGGAATGCCAGCGCTAGTGGTGTACAGAGAGAGTCAACAAAGCTGTGATCATGTTTCA 1282
Db      416 AspSerValThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHis 435
OY      1283 GACATCTTATCCCTCCCGCCCAATCAAGAGAGAGAGAGAGAGCTGTGCTCCGCGGCC 1342
Db      436 AspIlePheHisValIlePheProGlnThrHis--GluAspGluIleGluPheIlePheAla 454
OY      1343 AATGAAATGCAAGACGCGCTTCCCATTTGTACAAGTCAACGCGCGTTTAAATCCGAG 1402
Db      455 SerLysLysThrGlyThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSer 474
OY      1403 GGCATGATGGAGTGAAGCCCTTCAGCGCCGAGGAAGATGAATTAATGCGCCCATTAAG 1462
Db      475 LysTyrLysArgSerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLys 494
OY      1463 GAAGAGATGCTGTGACAGAGGTTGAATGGAGGTTTGGGAGAGACAGGCTCCAAAGATC 1522
Db      495 GluGluIleThrIleThrIleSerGlyLysIleProLysValLeuGluArgHisGlySerAsnIle 514
OY      1523 TGGGTCAATGAGAGAACAGTGGTGTACTTCCAGGAGCAACAAAGACAGCGCTGGAG 1582
Db      515 TyrValAspGluAlaArgLysLeuValTyrPheGluGluThrLysAspSerProLeuGlu 534
OY      1583 CACCAACCTTACAGTGGTACAGTATGAGCGCGCGCGAGATCGTACGCTCCACAGCGCC 1642
Db      535 HisHisLeuTyrValThrSerTyrAlaAsnProLysIleValValAlaGluLeuThrAspArg 554
OY      1643 GGCCTTCCCATAGCTGCTGCATGAGCGACAAATCTTCACATGTTGCGAGCATTAAGC 1702
Db      555 GluTyrGlnHisSerCysLysSerArgHisCysAspPheIleSerLysTyrSer 574
OY      1703 AGCGTGAAGACGCGCGCTGCTGACAGCTTACAAGCTGAAGCGCGCCGAGAGACAGCC 1762
Db      575 AsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspPro 594
OY      1763 CTGCACAAGAGCGCCCTTCTGGGCTAAGCATGATGAGAGGACGACGCTCCCGCGGAT 1822
:::|||||
:::|||||
:::|||||

```

```

Db      595 ValHisLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAsp 614
OY      1823 TATGTCCTCCAGAGATCTTCATTCATCCACAGCGGCTCGAGTGGCTTACGGCATG 1882
Db      615 TyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet 634
OY      1883 AATCAAGGCCCAAGGCTTCCAGCCAGAGGAAGAGACACCCAGCTCTTGTGTAT 1942
Db      635 LeuTyrLysProHisAspLeuGlnProGluLysLysTyrProThrValIlePheIleTyr 654
OY      1943 GGAAGCCCCAGAGTCAAGCTGGTGAATTAATCTTCAAAGCATCAAGTACTCGGCTC 2002
Db      655 GluLysProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 674
OY      2003 AACACATGCGCTCCCTCGGCTACGCGCTGGTGTGATGTCAGGAGAGGCTCTGTCAG 2062
Db      675 AsnThrLeuAlaSerLeuGluTyrValValValAlaValIleAspAsnArgGlySerCysHis 694
OY      2063 CGAGGCTTGGTTGGAAGGCGCCCTGAAAAACCAATGGCGCAGTGAAGATGAGAGAC 2122
Db      695 ArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspArg 714
OY      2123 CAGGTGAAGGCGCTCAAGTTCGTGGCCGAGAAATAGTATGCTTCATGACGTGAGCCGATT 2182
Db      715 GlnValGluGluLeuGlnTyrLeuAlaSerGlnTyrAspPheIleAspLeuAspArgVal 734
OY      2183 GCGATCCATGGCTGCTCAAGGCGGCTCCCTCGCTCATGAGGCGTAAATGCAAGCC 2242
Db      735 GlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSer 754
OY      2243 CAGGTGTTCAAGGTGGCCATCGCGGCTGCCCCGCTGCAACCTGTGATGCGCTTACAGACA 2302
Db      755 AspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThr 774
OY      2303 GCGTACACTGAAGCCCTACATGAGAGCTCCCTGAGAAACAACAGACGCGCTATGAGCGGCT 2362
Db      775 GlyTyrThrGlnArgTyrLysMetGlnLysProAspGlnAsnGluGlnTyrTyrLeuGly 794
OY      2363 TCCGTGGCCCTGCAAGTGAAGAGCTGCGCAATGAGGCGCAACGCTGTGATATCCAC 2422
Db      795 SerAlaIleMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHis 814
OY      2423 GCGTCCCTGACAGAAAGTGCACATTTTCCACAAATCTCTCGTCCCAATGATC 2482
Db      815 GlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuVal 834
OY      2483 CGACAGAGGAACCTTCACAGCTGACATGCCCAACGAGAGACAGATTCGCTGC 2542
Db      835 ArgAlaGluLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgVal 854
OY      2543 CCCGAGTGGGCGAGCACTATGAAGTACAGTGTGCTGCACTTCTTACAGGAATACCTC 2599
Db      855 ProLysSerGlyGlnHisTyrGluLeuHisValLeuLeuHisTyrLeuGlnGluAsnLeu 873

RESULT 3
O9HBM5
ID O9HBM5 PRELIMINARY; PRT; 882 AA.
AC O9HBM5:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-JUN-2002 (Tremblrel. 16, Last sequence update)
DE Dipeptidyl peptidase 8.
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA Gortelli M.D.;
RT "Cloning, expression and chromosomal localization of a novel human

```



```
|||||
Db 653 TyGtGlyProGlnValGlnLeuValAsnAsnArgPheUysGlyValIysTyrPheArg 672
OY 2000 CTCACACACATGGGCTCTGCGGTACCGCGGTGGTGGATTCAGGACGAGGCTCTGT 2059
Db 673 LeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCys 692
OY 2060 CAGCAGAGGCTTGGTTCGAAAGGGCCGTAAACCAATGGGCGAGGTGAGATCGAG 2119
Db 693 HisArgGlyLeuUysPheGlnGlyAlaPheUysTyrIleUysMetGlyGlnIleGluIleAsp 712
OY 2120 GACCAGGTGAGGGCCCTGCGATTCGTCGCCGAGAAATGATGGCTTCATCGACCTGAGCCGA 2179
Db 713 AspGlnValIleGlnLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 732
OY 2180 GTTGCCATGATGATGCTGCTGCTACAGGGGGGCTCTCTGCTGCTGATGGGGCTAATCCAG 2229
Db 733 ValGlyIleHisGlyTyrPserTyrGlyGlyTyrLeuSerIleUleMetAlaLeuMetGlnArg 752
OY 2240 CCCCAGGTGTCAGAGTGGCCATCGCGGGTGCCTCCAGCCGTGATGAGCCCTACGAC 2299
Db 753 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAsp 772
OY 2300 ACAGAGTACACTGAGCCCTACATGAGACGCTCCCTGAGAACACACGCGCTATGAGCGC 2359
Db 773 ThrGlyTyrThrGlnArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeu 792
OY 2360 GGTTCGCTGGCTCGACAGTGGAGAACGCTGCCAATGAGCCCAACCGCTGCTATGCTC 2419
Db 793 GlySerValAlaMetGlnAlaGlnLysPheProSerGlnProAsnArgLeuLeuLeu 812
OY 2420 CAGCGCTCTCTGAGCAAGAAAGTGCACCTTTTCCACAAACTCTCTGCTCCCACTG 2479
Db 813 HisGlyPheLeuAspGlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
OY 2480 ATCCGAGCAGAGGAACCTTACAGCTCCAGATCTACACCCCAAGAGACACACATTCGC 2539
Db 833 ValArgIleAlaLysProTyrAspLeuGlnIleTyrProGlnGlnAsnGlnSerIleArg 852
OY 2540 TGCCTCAGTGGGCGAGCACTAGATGACGCTGCTGCTGCTTACAGGAATCACTG 2599
Db 853 ValProGlnSerGlyLysIleHisTyrGlnLeuHisLeuLeuHisTyrLeuGlnGlnAsnLeu 872

RESULT 4
OY 075273 PRELIMINARY: PRT; 508 AA.
AC 075273:
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE R26984_1 (Fragment).
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamedin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Atlix C., Andrade T., Frankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Kromavishi B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carraro A.V.;
ST "Sequence analysis of a 2.5 Mb region in 19p13.3."
XL Submitted (SEP-1998) to the EMBL/Genbank/DBD databases.
DR EMBL; AC005594; AAC33801.1; -.
DR MEROPS; S09.019; -.
DR InterPro; IPR002410; Pro_amoPase.
DR InterPro; IPR000379; Ser_estrs_site.
DR PRINTS; PRO0793; PROAMOPTASE.
FT NON_TER 1 1
```

```
SQ SEQUENCE 508 AA; 57750 MW; 2F4D6645BE2D2C89 CRC64;
Alignment Scores:
Pred. No.: 3 36e-163 Length: 508
Score: 2572.00 Matches: 493
Percent Similarity: 91.34% Conservative: 3
Best Local Similarity: 90.79% Mismatches: 3
Query Match: 53.64% Indels: 44
DB: 4 Gaps: 5

US-09-976-674-4 (1-2617) x 075273 (1-508)
OY 998 ATGCTTCGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1057
Db 1 IleValSerThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 11
OY 1058 GAGTACATCCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1117
Db 21 GlnTyrIleAlaArgAlaGly-----AlaThrAlaMetPheLeu 33
OY 1118 GACCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177
Db 34 AspArgProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 53
OY 1178 ACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
Db 54 ThrGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 73
OY 1238 TATGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297
Db 74 TyrValValTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 93
OY 1298 TTCCTCCCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
Db 94 PheProGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 113
OY 1358 GCGCTTGCATTTGTACAAAGTACCGCGGTTTAAATCCAGAGGCTACGATTCGAGT 1417
Db 114 GlyPheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTyrPser 133
OY 1418 GAGCGCTTCAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1477
Db 134 GlnProPheSerProGlyGlnGly-----GlnGlnSerLeu 145
OY 1478 ACCAGCGGTGAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1537
Db 146 ThrAsnAla-----IleThrValAsnGlnGln 154
OY 1538 ACCAGAGCTGTGATCTTCAGAGGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
Db 155 ThrLysLeuValTyrPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 174
OY 1598 GTACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
Db 175 ValSerTyrGlnAlaAlaGlyGlnIleValAlaArgLeuThrThrProGlyPheSerHisSer 194
OY 1658 TGCCTCATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1717
Db 195 CysSerMetSerGlnAsnPheAspMetPheValSerHisTyrSerSerValSerThrPro 214
OY 1718 CCTGCTGACAGCTTACAGAGTGAAGTGAAGGCGCGCGAGAGAGAGAGAGAGAGAGAGAG 1777
Db 215 ProCysValHisValTyrLysLeuSerGlyProAspAspAspProLeuHisLysGlnPro 234
OY 1778 CGCTTCGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1837
Db 235 ArgPheThrAlaSerMetCysGlnAlaAla-----Lys 245
OY 1838 ATCTTCATTTCCACAGCGCTGAGATGTCGGCTACAGGACATGATCTCAAGAGAGAGAG 1897
Db 246 IlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMetIleTyrGlyProHis 265
OY 1898 GCCTTCAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
```

```

|||||
Db 266 AlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyrGlyGlyProGlnVal 285
QY 1958 CAGCTGGGATACCTCTTAAAGGATCAAGTACTGGGGCTCACACACTGGCCCTCC 2017
Db 286 GlnPheValAsnAsnSerPheLysGlyLleLysTyrLeuArgLeuAsnThrLeuAlaSer 305
QY 2018 CTGGGCTACGGCGTGTGATGATGACGGCGGGCTCTCTGACGGAGGGCTTCGGGTC 2077
Db 306 LeuGlyTyrAlaValValValIleAspGlyArgGlySerGlyGlnArgGlyLeuArgPhe 325
QY 2078 GAAGGGGCGCTGAAAAACCAATGCGCCAGGTGAGATCGAGACGACGAGTGGAGCGCTG 2137
Db 326 GlnGlyAlaLeuLysAsnGlnMetGlyGlnValGlnIleGlnAspGlnValGlnGlyLeu 345
QY 2138 CAGTTCGTGGCGGAGAAAGTATGAGCTTATCATGACCTGAGCCGAGTGGCATCAGGCGG 2197
Db 346 GlnPheValAlaGlnLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTyr 365
QY 2198 TCCTACGGGGGCTCTCTCTGCTCATGGGGCTAATCCAGACGCCAGGTGTCAAGGTG 2257
Db 366 SerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysVal 385
QY 2258 GCCATCGCGGGTCCCGGTCACCGTGTGATGGCTTACGACAGACAGGGTACACTGAGCGC 2317
Db 386 AlaIleAlaGlyAlaProValThrValTyrPheAlaTyrAspThrGlnGlyTyrThrGlnArg 405
QY 2318 TACATGACAGCTCCCTGAGACACACGACGAGCTATGAGCGGGGTCCGCGCCCTGCAC 2377
Db 406 TyrMetLysPValProGlnAsnAsnGlnHisGlyTyrGlnAlaGlySerValAlaIleHis 425
QY 2378 GTGAGAGAAGCTGCCCAATGAGCCCAACCGCTGTATCCCTCAAGGGTCTCGAGCGAA 2437
Db 426 ValGlnLysLeuProAsnGlnProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGln 445
QY 2438 AACGTGACATTTTCCACACAAACTCTCTCTGCCAAGTATCCGAGCAGGAGAAACT 2497
Db 446 AsnValHisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysPro 465
QY 2498 TACCAAGCTC-----CAGATCTACCCCAACGAGAGACAC 2530
Db 466 TyrIleGlnGlnValAlaLeuProProValSerProGlnIleTyrProAsnGlnArgHis 485
QY 2531 AGATATGCTGCTCCCGAGTGGGCGGAGACATATGAGTACAGTTCCTGCTACATTTCTACAG 2590
Db 486 SerIleArgCysProGlnSerGlyGlnHisTyrGlnValIlePheLeuHisPheLeuGln 505
QY 2591 GAATACCTC 2599
Db 506 GlnTyrLeu 508

RESULT 5
Q9BVR3 PRELIMINARY: PRT: 439 AA.
ID Q9BVR3;
AC Q9BVR3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 49.9 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC000970: AAH00970.1: -
DR MEROPS: S09.019: -
DR InterPro: IPR001375: Peptidase_S9.
DR InterPro: IPR000379: Ser_estrs_site.
DR Pfam: PF00326: Peptidase_S9; 1.

```

```

KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 439 AA; 49926 MW; A18BBA9E12092BAF CRC64;

Alignment Scores:
Pred. No.: 1,28e-150
Score: 2383.00
Percent Similarity: 100.00%
Best local Similarity: 100.00%
Query Match: 49.70%
DB: 4
Gaps: 0

US-09-976-674-4 (1-2617) x Q9BVR3 (1-439)
QY 1283 GACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGACGAGCTCTGCTTCTCCGCC 1342
Db 1 AspllePheTyrProPheProGlnSerGlnGlyGlnAspGlnLeuCysPheLeuArgAla 20
QY 1343 AATGAATCAGACCGGCTTCCCATTTGTACAAAGTACCGCGGCTTTAAATCCAG 1402
Db 21 AsnGlyCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSerGln 40
QY 1403 GCTACGATTCGAGTGAAGCCCTTCAGCCCGGGGAAGATGAATTAAGTCCCATTAAG 1462
Db 41 GlyTyrAspTyrSerGlnProPheSerProGlyGlnAspGlnPheLysCysProIleLys 60
QY 1463 GAAGAGATTCGCTGACACGCGGTGAATGGAGGTTTGGCGAGGACGCGCTCCAAATC 1522
Db 61 GlnGlnIleAlaLeuThrSerGlyGlnTyrPheValLeuAlaArgHisGlySerLysIle 80
QY 1523 TGGGTCATGAGAGAGACCAACAGCTGCTGACTCCAGGGCAGACAGGACGCGCTGGAG 1582
Db 81 TrpValAsnGlnGlnThrLysLeuValTyrPheGlnGlyThrLysAspThrProLeuGln 100
QY 1583 CACCACTCTACGTGTCAGCTATGAGCGCGCGCGAGATCGTACGCGCTCACACGCC 1642
Db 101 HisHisLeuTyrValIleValSerTyrGlnAlaIleGlyGlnIleValArgLeuThrThrPro 120
QY 1643 GCGTCTCCCATAGCTGCTCCATGACGCGGATTCGACATGTTCTGATGACCATACAGC 1702
Db 121 GlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheValSerHisLysSer 140
QY 1703 AGCGTGACGACGCGCGCTGCGTGCAGCTTACAAAGTACGAGCGCGCGCGAGACGCC 1762
Db 141 SerValSerThrProProCysValHisValTyrLysLeuSerGlyProAspAspPro 160
QY 1763 CTGCACAAGACGCCCGCTTCTGGCTAGCATGATGAGGACGACCGCTGCCCGGAT 1822
Db 161 LeuHisLysGlnProArgPheThrPheAlaSerMetGlnAlaIleAspCysProProAsp 180
QY 1823 TATGTTCTCCAGAGATCTTCCATTTCCACACGCGCTGGATGTCGGCTCTAGGCGATG 1882
Db 181 TyrValProProGlnIlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMet 200
QY 1883 ATCTCAAGCCCCACGCGCTTCGAGCGGAGGAAGACGCCACCGCTCTTGTATAT 1942
Db 201 IleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyr 220
QY 1943 GGAGGCCCGCAGGTGACCTGGTGAATACCTCTTCAAGGACATCACTACTTGGCGCTC 2002
Db 221 GlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyLleLysTyrLeuArgLeu 240
QY 2003 AACACACTGGGCTCCCTGGGCTAGCGCGCTGGTGTGATGACGCGAGGGGCTCTGTAG 2062
Db 241 AsnThrLeuAlaSerLeuGlyTyrAlaValAlaValIleAspGlyArgGlySerCysGln 260
QY 2063 CGAGGGCTTCGTTGCAAGGGGCGCTGAAAAACCAATGGCGCAGGTGCAATCGAGAC 2122
Db 261 ArgGlyLeuAlaArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnValGlnIleGlnAsp 280
QY 2123 CAGGTGAGGGCGCTGCACTGCTGTCGCGCGAGAGATAGCTTCATCGACCTGAGCGAGTT 2182
Db 281 GlnValGlnGlyLeuGlnPheValAlaGlnLysTyrGlyPheIleAspLeuSerArgVal 300

```


Db 379 ProHiscysValserLeuTyrlsLeuSerProGluAspProThrcysLysThr 398
 QY 1775 CCCCCTCTGGGCTAGATGAGACCCAGCTCCCGCGATTAATGTTCTCCA 1834
 Db 399 LysGluPheThrPalatThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 418
 QY 1835 GAGATCTTCCATTTCCACACAGCGCGTGGCGGTCTCCAGGCAATCTACAAAGCC 1894
 Db 419 GluIlePheSerPheGluSerThrThrIlePheThrLeuTyrlsGlyMetLeuTyrlsPro 438
 QY 1895 CACGCGCTTGCAGCCAGGAGAACACCCAGCTCTTGTATATGAGGCCCGCCAG 1954
 Db 439 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyProGln 458
 QY 1955 GTGCAGCTGGTGAATAACTCTCTCAAGCATCAAGTATCGCGCTCAACACACTGGCC 2014
 Db 459 ValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrIleuAla 478
 QY 2015 TCCCTGGGCTACGCGCGTGTGTGATGATGAGCGAGGGCTCCTGACAGAGGCTTGG 2074
 Db 479 SerLeuGlyTyrValValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLys 498
 QY 2075 TTGCAAGGGGCGCTGAAAAACCAATGGCCAGGTGAGATCGAGACCAAGTGGAGGGC 2134
 Db 499 PheGluGlyAlaPheLysTyrLysMet----- 507
 QY 2135 CTGCAGTTCGTGGCCGAGAAAGATATGGCTTCATCGACTGAGCCGAGTTGCCATCCATGCC 2194
 Db 507 ----- 507
 QY 2195 TGGTCTACGGGGGCTCTCTCGCTCATGGGGCTAATCCACAGCCCGAGGTGTTCAAG 2254
 Db 507 ----- 507
 QY 2255 GTGGCCATCGGGGTCGCCCGGTACCGCTGTGATGGCTACGACAGAGGTACACTGAG 2314
 Db 508 ValAlaIleAlaGlyAlaIleProValThrLeuThrPheTyrAspThrGlyTyrThrGln 527
 QY 2315 CGCTACATGAGAGTCCCTGAGAACACCAAGCAGCAGCTATGAGGGCGGTTCCGTGGCCCTG 2374
 Db 528 ArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrLysGlySerValAlaMet 547
 QY 2375 CACGTGAGAGAGCTGCCAATGAGCCCAACCGCTGTATTCCTCCAGGGCTCTCTGGAC 2434
 Db 548 GlnAlaGlnLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 567
 QY 2435 GAAAAAGTGAAGCTTTTCCACACAAACTCTCTGATCCCAACTGATCCGAGCAGGAAA 2494
 Db 568 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 587
 QY 2495 CCTTACAGCTCCAGATGATACCCCAAGAGAGACAGACAGTATTCGTCGCCCGAGTGGGC 2554
 Db 588 ProTyrAspLeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGluSerGly 607
 QY 2555 GAGCAGTATGAAGTCAAGTGTGCTGCTTCTACAGAGATACCTC 2599
 Db 608 GlnHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlnAsnLeu 622
 RESULT 7
 075868 PRELIMINARY: PRT: 432 AA.
 AC 075868:
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE R31083.1 (Fragment).
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 CX NCBI_TaxID=9606;
 RN 11)
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,

RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 RA Dargahan L., Ertler A., Christensen M., Georgescu A., Avila J., Liu S.,
 RA Atlix C., Andreise T., Frankheim M., Amico-Keller G., Coffield J.,
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
 RA Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RA Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of a 2.5 kb region in 19p13.3";
 RL Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AC005783; AAC62840.1;
 FT NON_TER 432
 FT 432
 SQ SEQUENCE 432 AA; 48595 MW; 64E2B85BE0523A7E CRC64;
 Alignment Scores:
 Pred. No.: 1.07e-115 Length: 432
 Score: 1859.50 Matches: 354
 Percent Similarity: 97.52% Conservative: 0
 Best Local Similarity: 97.52% Mismatches: 0
 Query Match: 38.78% Indels: 9
 DB: 4 Gaps: 2
 US-09-976-674-4 (1-2617) x 075868 (1-432)
 QY 11 ATGGCCACACCGGGAGCCCAACGCGCAGCGAGCGAGCGCCAGATGACCG 70
 Db 72 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspPro 91
 QY 71 GCCGCCGCTTCCAGGTCCAGAACCACTCGTGGAGCGGCTCCGAGCATCCACGCG 130
 Db 92 AlaAlaArgPheGlnValGlnLysHisSerTyrAspGlyLeuArgSerIleIleHisGly 111
 QY 131 AGCCCAAGTACTGGGGCTCATTTGTCAACAAGCGCCCGACGATTCAGTTTGTGAG 190
 Db 112 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 131
 QY 191 AAGAGATGAGTGGGGCCCGACATCCCGCCCTCTACTGCTGAGTGGAGTGCATATGAG 250
 Db 132 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 151
 QY 251 AGCCGAGAACTCCCTCTCTACTCTGATGATCCCAAGAGTCCGGAAGAGCTCTG 310
 Db 152 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysValArgLysGluAlaLeu 171
 QY 311 CTGCTCTCTGCTGGAAACAGATGCTGATCTTCAGGCCACCGCCCGACCATGGGTC 370
 Db 172 LeuLeuLeuSerTyrPheGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 191
 QY 371 TACTCTCGAGAGAGAGCTGCTGAGGAGCGGAAAGCGCTGGGCTCTGGCATCAC 430
 Db 192 TyrSerArgGluGluGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr 211
 QY 431 TCTTACAGCTTCCACAGGAGAGTGGCTCTCTCTTCCAGGCCAGACAGCTCTTC 490
 Db 212 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 231
 QY 491 CACTGTCCGAGCGGGCGGAAGAAGCGCTTCATG-----GTGTCC 529
 Db 232 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProGlyProGlyCysValSer 251
 QY 530 CCTATGAACCGCTGGAATGATCAAGACCCAGTGCCTGAGGCGCCCGATGAGCCCAAAATC 589
 Db 252 PromethLysProLeuGlnIleLysThrGlnCysSerGlyProArgMetAspProLysIle 271
 QY 590 TGCCTGCGGACCGCTGCTTCTCTCTTCAATTAACAGAGCTGTGGGTGGCCAAAC 649
 Db 272 CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTyrValAlaAsn 291
 QY 650 ATCGAGAGAGCGGAGAGCGGCGGTGACCTTCTCCACCAAGTTTATCCAAATGCTCTG 709
 Db 292 IleGlnThrGlyGlnGlnArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 311
 QY 710 GATGACCCCAAGTGTGGGGGTGGCCACTCCGATACAGGAAGAGTGTGACCGCTTC 769


```
Dh 312 AspaSProlYSSerAlaGlyValAlaThrPheValIleGlnGluGluPheAspArgPhe 331
QY 770 ACTGGTACTGTGGTGGCCACACGCTCTGGAGAGGTTCCAGAGGCTTCAGACGCTG 829
Dh 332 ThrIlyTrpTrpCysProThAlaSerTrpGlu-----GluGlyLeuLysThrIleu 349
QY 830 CGAATCTGTATGAGGAGAGTGCATGATCCGAGGTGAGGTGCATTCACGTCCTTCCT 889
Dh 350 ArgIleuTrpGluGluValAspGluSerGluValGluValIleHisValProSerPro 369
QY 890 GCGGTAGAGAAAGAGAGAGCGACTCGTATGCGTATCCCGACAGAGAGAGAAATCCC 949
Dh 370 AlaIeuGluGluValGlyThrAspSerTrpArgTrpProArgTrpGlySerLysAsnPro 389
QY 950 AAGATTCCTTGAAGACTGGCTGATCCAGACTACAGACCGAGGAGAGATGCTCGAC 1009
Dh 390 LysIleAlaIeuLysIleAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThr 409
QY 1010 CAGGAGAGAGAGCTGGTGCACGCTTCACGCTGCTTCCCGAAGTGGAGTACATGCC 1069
Dh 410 GlnGluLysGluIleuValGlnProPheSerSerLeuPheProLysValGluTrpIleAla 429
QY 1070 AGGCGCGAG 1078
Dh 430 ArgAlaGly 432

RESULT 8
096NT8 PRELIMINARY; PRT; 312 AA.
ID 096NT8;
AC 096NT8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CNA FLJ30094 f1s, clone BNGH4100034, weakly similar to dipeptidyl
DE peptidase IV (EC 3.4.14.5).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kikunishi H., Kanda K., Matsuo K., Nakamura Y., Sekine M.,
RA Takahashi-Fujii A., Oshima A., Sugiyama K., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054656; BAB70784.1;
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; Ser_estr_site.
DR Pfam: PF00326; Peptidase_S9.1.
SQ SEQUENCE 312 AA; 35518 MW; ABE940AFC587717 CRC64;
```

Alignment Scores:

```
Pred. No.: 1,88e-103 Length: 312
Score: 1675.00 Matches: 311
Percent Similarity: 99.68% Conservative: 0
Best Local Similarity: 99.68% Mismatches: 1
Query Match: 34.93% Indels: 0
DB: 4 Gaps: 0
```

US-09-976-674-4 (1-2617) x 096NT8 (1-312)

```
QY 1664 ATGAGCCAGAACTTCGACATGTTGTGACGCACTACAGACGCTGACAGCGCCCTGC 1723
Dh 1 MetSerIlnsPheAspMetPheValSerHisTrpSerSerValSerThrProProCys 20
QY 1724 GTGCAGCTACAAAGCGAGCGGCGCGAGAGAGACCCCTGCACAGACGCGCCGCTTC 1783
Dh 21 ValHisValTrpLysLeuSerGlyProAspAspAspProLeuHisLysGlnProArgPhe 40
```

```
QY 1784 TGGCGTAGCATGATGAGAGGAGCCAGCTGCCCGCCGATGATGTTCCAGAGATCTTC 1843
Dh 41 TrpIaSerMetMetGluAlaIaIaSerCysProProAspArgValAlaProProGluIlePhe 60
QY 1844 CATTTCCACACGCGCTCGATGTCGGCTCTACGGCATGATTAACAAGCCCGCTTCG 1903
Dh 61 HisPheHisThrArgSerAspValArgLeuTrpGlyMetIleTrpLysProHisAlaIeu 80
QY 1904 CAGCGAGGAAAGAGACCCCGCTCTTTGATATGAGGCGCCCGAGTGCACCTG 1963
Dh 81 GluProGlyLysLysHisProThrValLeuPheValTrpGlyLysProGlnValGlnLeu 100
QY 1964 GTGAATTAATCTCTTCAAGGCAATCAAGTACTTGGCTGCGCTCAACACACTGCTCCCTGGC 2023
Dh 101 ValAsnAsnSerPheLysGlyIleLysTrpLeuArgLeuAsnTrpIleuAlaSerLeuGly 120
QY 2024 TACCGCGTGTGTGATTACGCGAGGGCTCTGTACGAGAGGGCTTCGGTTGGAAGG 2083
Dh 121 TyrAlaValAlaValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGly 140
QY 2084 GCGCTGAAAAACCAATGGGCGAGTGGAGATGAGAGGAGGAGGAGGAGGCGCTGCAGTTG 2143
Dh 141 AlaLeuLysAsnGlnMetGlyGlnValGlnIleLysPheValAlaIleHisGlyLeuGlnPhe 160
QY 2144 GTGGCCGAGAAATGATGCTTCATGACCTGACGCGAGCTTCGATGCTGATGCTGCTTAC 2203
Dh 161 ValAlaGluLysTrpGlyPheIleAspLeuSerArgValAlaIleHisGlyTrpSerTrp 180
QY 2204 GGGGCGTCTCTCTGCTCATGAGGGCTAATCCACAGCCCGAGGTTTCAAGTGGCCATC 2263
Dh 181 GlyLysPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIle 200
QY 2264 GCGGCGTCCCGCGTACCGCTGGATGGCTACGACGAGGATACATGAGCGCTACATG 2323
Dh 201 AlaCysAlaProValTrpValTrpMetAlaArgTrpArgTrpGlyTrpGluArgTrpMet 220
QY 2324 GACGTCCCTGAGAAACAACGACGACGAGCTATGAGGCGGGTTCGCTGGCCCTGCAGCTGGAG 2383
Dh 221 AspValProGluAsnAsnGlnHisGlyTrpGluAlaGlySerValAlaLeuHisValGlu 240
QY 2384 AAGTGGCCCAATGAGCCCAACGCGCTTATCTCTCCAGCGCTTCTCGAGAAAGCTG 2443
Dh 241 LysLeuProAsnGlnProAsnArgLeuLeuIleuHisGlyPheLeuAspGluAsnVal 260
QY 2444 CACTTTTCCACAAACTCTCTGCTCCCACTGATCCGAGAGGGAACCTTACAG 2503
Dh 261 HisPheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTrpGln 280
QY 2504 CTCGAGATCAACCCCAAGAGAGACAGATATTCGCTGCCCGAGTGGCGAGCACTAT 2563
Dh 281 LeuGlnIleTrpProAsnGlnAlaGlnHisSerTrpArgLysProGluSerGlyGlnHisTrp 300
QY 2564 GAAGTCAAGTTCGCTGCACTTTTACAGAGATACCTC 2599
Dh 301 GluValTrpLeuLeuHisPheLeuGlnGlyTrpLeu 312

RESULT 9
09VC20 PRELIMINARY; PRT; 1042 AA.
ID 09VC20;
AC 09VC20;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CG3744 protein.
GN CG3744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
```


RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blatz R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G.G., Nelson C.R., Miklos G.L.G.,
RA Abdl J.F., Asgharali A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baas A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernick B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke S., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Hock J.,
RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jostali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Munkulov G., Milshina N.V., Mobarry C., Morris J., Moshier A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splet E., Spradling A.C., Stapleton M., Stong R., Sun E.,
RA Svirskas R., Teeter C., Turner C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye Y., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003749; AF56537.1; --
DR MEROPS: 509.016; --
DR FlyBase: FBgn0039240; CG3744.
DR InterPro: IPR002469; DPriv_N_term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; SerEsters_site.
DR Pfam: PF00930; DPriv_N_term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
SQ SEQUENCE 1042 AA; 116706 MW; 1. B9J15DC365DF937 CRC64;

Alignment Scores:			
Pred. No.:	2,56e-100	Length:	1042
Score:	1631.50	Matches:	372
Percent Similarity:	54.76%	Conservative:	146
Best Local Similarity:	39.32%	Mismatches:	320
Query Match:	34.03%	Indels:	109
DB:	5	Gaps:	20

US-09-976-674-4 (1-2617) x Q9VC20 (1-1042)			
QY	5	CTTACCATGCGCCACCACCGGAGCCCA-----ACGGCCGACCGAGCGGACGCA	52
		: :	
Db	120	LeuSerSerProthrSerglyThrProProHisGlyLeuAspValAspGluGlyAspAsp	139
QY	53	GCCGCG-----ACAGATGACCCGCGC-----GCCCGC	79
Db	140	GlucyAspAspCysGluGluAspGluAspAspProValAspAsnAspGlyHisIleIle	159
QY	80	TTCACGATGCGAAGACACTCTGTGGAGAGGGCTCCGGAGACATCATCCACGGCACCAGCAG	139
		:::	
Db	160	AlaProthrProAsnLysSerThrPAlaGluAsnLysIleValGluGluIleArgLys	179
QY	140	TACTCGGGCTCATGTTCATACACAGCGCCGCCACGACTTCCAGTTTGTCGACGAGACGAGAT	199

Db	180	LysMetCysAsnLeuSerSerMetValProThiAsnValGlnProHrighiSteuSerAsp	199
Qy	200	GAGTCTGGGGCCCACTCCACCCGCTTCTACTACCTGGAGATCCATATGACGCGAG	259
Db	200	GlyArgAla-----ArgCysTyrPheLeuGlyTyrProProGlnSerTrpGlu	215
Qy	260	AACTCCCTCTACTCTGAGATTCGCCAGAGAGGTCCGGAAGAGGCTCTGCTCTC	316
Db	216	ThrThrLeuPheAlaAspIleAsnLeuThrGlnSerGluGlnGlnLeuVal	235
Qy	316	-----	316
Db	236	GlnArgLeuGlnGlyIleAlaSerAspGluTrpSerProThiMetAsnAlaGlySerPro	255
Qy	317	-----CTGCC	322
Db	256	ThrSerSerGlyHisGlnProAlaPheLeuPheAsnSerLeuProArgProArgLeuPro	275
Qy	323	TGGAAGCAGATGCTGATCAT---TTCCAGGCCACGCCCCACCATAGG-----	367
Db	276	TrpSerProLeuLeuGlnGlnProIleGlnSerSerGlyGlySerGlyValSerGlySer	295
Qy	368	-----GTCTACTCTGGGAGAGAGAGACTGCTGAGGAGCGGAACCCCTGGGGCTTTC	421
Db	296	AlaSerProTyrAlaArgGluTyrGlnLeuLeuGlnGluArgGlySerLeuSerThrTrp	315
Qy	422	GGCATCAACCTCCCTCAAGCTTCCACAGAGAGAGTGGCTCTCTCTCCAGGCCAGAAC	481
Db	316	GlyIleThrSerTyrGlnLeuHisLysProSerGlyLysLeuValPheProCysPheAsn	335
Qy	482	AGCCTCTCCACTCTCCGACGGCGGAGAACAGGCTTCATGGATGGTCCCTATGAAACCG	541
Db	336	AspLeuTyrGlnCysLeuAspThrGlyTyrAsnSerGlyLeu-----LeuPhePro	352
Qy	542	CTGGAATCAAGACCCAGTCTCAGGAGCCCGGATGACCCAAATCTGCCCTCCGAC	601
Db	353	ThrGlnLeuArgThrCysProGlnTrpThrAlaLeuAspProGlnIleCysProGlnAsn	372
Qy	602	CTGCTCTCTCTCTCTCATCATTAACAGAGGACTCTGGGTGGCCAACTAGACAGACG	661
Db	373	SerAspMetIleAlaTyrIleSerAspCysAspLeuPheValThrHisLeuSerGly	392
Qy	662	GAGAGCGGCGGCTGACCTTCTGCCACCAAGGTTTATCCATGCTCTGGATGACCCCAAG	721
Db	393	HisGluLysArgLeuThrTyrThrSerThrGlyArgHisSerTyrValAspAspAlaLeu	412
Qy	722	TCGCGGGGTGGCCACCTTGTGCATACAGGAAGATTGCACCGTCTCACTGGATGCG	781
Db	413	SerAlaGlyValProSerTyrValMetGlnGluPheSerArgTyrGlnGlyPheTrp	432
Qy	782	TGGTCCCCACAGGCTCTCTGGGAAGGTTCAGAGGGCTCAACAGCTGGGAATCTGTAT	841
Db	433	TrpGlnPro-----HisSerAsnAspGlyIle-----TyrArgIleValTyr	446
Qy	842	GAGAGATCGATGATCGGAGGTGAGGATATTCACGATGCTCCCTGCTCGCTAGAACAA	901
Db	447	GlnGluValAspGlnSerGluValSerValTyrThrPheProSerSerThrAlaMetHis	466
Qy	902	AGGAAGACGACTGTATGGTATGCCACCCAGAGACAGGACGAAGAATCCAAAGATTGCC	961
Db	467	GlyArgValAspGluTyrArgPheProArgThrGlySerProAsnAlaLysSerLysLeu	486
Qy	962	AAACTGGCTAGTTCGCACTGACAGCCAGGCAACAATGCTCTCGACCCAGAGAGAGAG	1021
Db	487	LysLeuValGlnPheValLeuAsnGlnAlaLeuGlnValSerGluIleAlaIleLysAsp	506
Qy	1022	CTGGAGACCTTCACCTGCTGCTGTGTCCGGAAGGTGAGATCGCCAGGGCCGGGAGG	1081
Db	507	LeuProTyrSerLeuLeuAlaValPheSerTrpLeuGluTyrIleValArgValGlyTrp	526
Qy	1082	ACCGGGATGGCAATACAGGCTGGGGCATTTCTTGAGCGGCGCCAGAGAGTGGCTCAG	1141

Dh 527 ThrProAspAlaLysTyrValTTrpValGInGlyLeuAspArgLysGInGlnArgLeuAsp 546
Qy 1142 CTCGTCTCTCCCTCCCGGCGCTGTTCATCCCGAGACAGAGATGAGAGACCGGCTA 1201
Dh 547 ValIleuLeuIleProLeuAspAsnPhcysGluSerTyrSerSerGInValSerThrPro 566
Qy 1202 GCGCTGCGC-----AGAGCTGTCCCGAGAAATGTCACGCG-- 1237
Dh 567 ThrAspSerIleGlyAspAsnIleSerTrpArgSerLeuTyrSerArgThrIleThrProLeu 586
Qy 1238 TATGTGGTGAAG 1257
Dh 587 GlnAlaValLeuTyrThrGlnArgSerSerTrpIleAsnValHisAspMetLeuHisPhe 606
Qy 1298 TTCCCAATACAG 1357
Dh 607 LeuAspLeuThr-----GluThrSerValThrPheLeuThrAlaSerGlu---GluThr 623
Qy 1358 GCGCTGTGCGCTTTGTACAAAGTACCGCGC--GTTTAAATCCAGAGCGCTACGATTGG 1414
Dh 624 GlyPheArgHisLeuTyrLeuValThrAlaSerLeuLeuSerGlnAlaAsnGlyGln 643
Qy 1415 AGTAGCGCTTCAGCGCGCGGAGAA-----GATGATTTAAGTGC 1453
Dh 644 ProAspProGlySerValGlyAlaGlnProSerPheValAspLeuSerAlaLeuGlnPro 663
Qy 1454 CCCATTAGAAGAGATTCGTCTGACAGCGGTGATGGAGAGTTTGGCGAGCGACGGC 1513
Dh 664 ArgIleLeuAsnLysValAlaLeuThrSerGlyGlnTrpGluValLeuAlaArgAsn--- 682
Qy 1514 TCCAAGATTCGGGTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1573
Dh 683 -----LeuTrpValAspLysAlaAsnLysLeuValTyrPheValGlyLeuArgAspThr 700
Qy 1574 CCGGTGAGAGACCGACCTCTGATGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGCTC 1633
Dh 701 ProLeuGlnLysHisLeuTyrValValSerLeuGlnArgProGlnHisIleArgLeuLeu 720
Qy 1634 ACCAGCGCGCGCTTCCTCCATAGCTGCTCCATGAGCCAGAACTTCGACATGTCGTGACG 1693
Dh 721 ThrGluProGlyTyrSerTyrLeuValGlnPheAspArgLysLysLeuMetLeuLeu 740
Qy 1694 CACTACAGACGCTGAGACAGCGCGCGCTGCGTGCACGCTTACAGAGCTGAGCGCGCGGAC 1753
Dh 741 ValTyrGlyAsnIleGln-ArgLeuProSerCysLys-----ValMetArgVal 756
Qy 1754 GAGGACCGCGCGACAGACCGCGCGCTTCGCGCTGACG-----AT 1794
Dh 756 IAsnGlnThrCysSerAsnGlyValAsnGlyIleGlnIleSerLeuValGlyTyrLe 776
Qy 1795 GATGAGAGAGACCGACGCGCGCGATTTATGTTCTTCAGAGATCTTCATTTCCACAC 1854
Dh 776 uHisGlnGlyGlyLysProGluProGlnTyrCys---ProGlnIlePheSerProGlnLe 795
Qy 1855 GCGGTGCGATGTGGCGCTTACAGCATGATCTACAGCCCGACCGCTTCGACGCGAGGAA 1914
Dh 795 uProSerGlyAspIleValAlaLysMetValPheLysProHisAsnPheGluLeuGlyVal 815
Qy 1915 GAAGGACCGCGCGCTCTTTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974
Dh 815 LysTyrTrpThrValLeuAsnValTyrGlyGlyProGluValGlnThrValAsnAsnThr 835
Qy 1975 CTTCAAAGGATCAAGTACTTGGCGCTCAACACACATGCGCTCCCTGGGCTACGCGCTGGT 2034
Dh 835 rPheLysGlyLysHisGlnLeuArgMetHisMetLeuAlaIleGlnGlyTyrCysValIle 855
Qy 2035 TGTGATTGACGAGCGGCGCTCTGTACAGCGAGCGGCTTCGTTGACAGCGCGCGTGA 2094
Dh 855 ecysIleAspSerArgLysArgHisArgGlyLysArgPheGluSerHisIleArgGly 875
Qy 2095 CCAATGGCGAGGTGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
Dh 875 yArgMetGlyGlnValGlnLeuThrAspGlnValAspAlaLeuArgSerLeuSerAspGly 895

Qy 2155 GATAGCTTCATCGACCTGAGCGAGTTCGCCATTCATGCGTGGTCTTACGGGGCTTCCT 2214
Dh 895 nLeuGlyTyrIleAspMetAspArgValAlaIleHisGlyTrpSerTyrGlyGlyTyrLe 915
Qy 2215 CTCGTGATGGGGCTAAATCCAAAGAGCCCGACGCTTTCAGGTGGCCATCCGGGGTCCCG 2274
Dh 915 uSerLeuMetGlyLeuValGlnTyrProLysIlePheLysValAlaIleAlaGlyAlaP 935
Qy 2275 GGTACCGCTGATGGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2334
Dh 935 oValThrAsnTrpGluTyrTrpAspTrpGlyTyrThrGlnArgTyrMetAspMetProG 955
Qy 2335 GAACACACAGACCGCTGTATGAGCGGGGTTCGCGCGCTGACGAGAGAGAGAGAGAGAG 2394
Dh 955 nAsnAsnGlnAlaGlyTyrSerAlaGlySerValLeuGlnTyrValAsnSerPheProG 975
Qy 2395 TGAAGCCCAACCGCTTGTATTCCTTCACAGCGCTTCCTGACGAGAAAGCTGACTTTTCA 2454
Dh 975 uGluAspLysArgLeuLeuLeuIleHisGlyLeuIleAspGluAsnValHisPheCysH 995
Qy 2455 CACAACCTCTGCTGCTCCCACTGATCCGAGCGAGGAGAACTTACACAGCGTACGATCTA 2514
Dh 995 sThrSerArgLeuIleSerAlaLeuAsnLysAlaAsnLysPheTrpGlyGlnHisLeuPh 1015
Qy 2515 CCCCAACAGAGACAGATATTCGCTCCCGAGCTCGGGCGGACATATGATGACGTT 2574
Dh 1015 eProGluGlnArgHisSerLeuArgAsnLeuGlnSerAsnLysAsnTyrGlnThrLysLe 1035
Qy 2575 GCTGCACTTTCATACAG 2590
Dh 1035 uLeuSerPheLeuGln 1040
RESULT 10
09VC19
ID 09VC19 PRELIMINARY: PRT: 1102 AA.
AC 09VC19;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CG3744 protein.
GN CG3744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Ramanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Rao K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Ahril J.F., Aghayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
Bailow R.M., Besu A.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
Rao K.H., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,


```
Dh 704 ProAspProGlySerValGlyAlaIleProSerPheValAspLeuSerAlaLeuGlnPro 723
Qy 1454 CCCATTAAAGAGAGATTGCTGTGACAGCGGTGAATGGAGAGTTTGGCGAGCGACGGC 1513
Dh 724 ArgIleuAsnLysValAlaIleuThrSerGlyGluTrpGluValLeuAlaArgAsn--- 742
Qy 1514 TCACAAGATTGGGGTCAATAGAGAGACCAAGCTGGTGTACTCTCCAGGCGACCAAGACACG 1573
Dh 743 -----LeuTrpValAspLysAlaAsnLysLeuValTyrPheValGlyLeuArgAspThr 760
Qy 1574 CCGGAGCGACACAGCTGTAGGTGATGATGAGCGGGCGCGGAGATGTAGCGCTC 1633
Dh 761 ProLeuGlnLysHisLeuTyrValAlaSerLeuGlnLysProGlnHisIleArgLeuLeu 780
Qy 1634 AACAGCGCGGGCTTCCATGATGCTGTCCATGAGCCAGAACTTCAGATGTTCCTGACG 1693
Dh 781 ThrGluProGlyTyrSerTyrLeuValGluPheAspArgLysCysLysLeuMetLeuLeu 800
Qy 1694 CACTACAGACGGGTGACAGACCGCGCTGCGTGCACGTCTACAGCTGAGCGGCGCCGAC 1753
Dh 801 ValTyrCysAsnIleGln-ArgLeuProSerCysLys-----ValMetArgVal 816
Qy 1754 GACGACCCCGTGCACAGACGCGCGCTTCGGGCTAGC-----AT 1794
Dh 816 IAsnGlnThrCysSerAsnGlyValAlaAsnGlyIleGlnIleSerLeuValGlyTyrLe 836
Qy 1795 GATGAGAGCAGCAGCAGCTCCCGCGATTAATGTTCTCCAGATCTTCATTTCCACAC 1854
Dh 836 uHisGluGlyLysProGluProGlnTyrCys---ProGlnIlePheSerProGlnIle 855
Qy 1855 GCGCTGGATGCGCGCTCTACGCGATGATACAGCCCGCAGCGCTTCAGACCGGAA 1914
Dh 855 uProSerGlyAspIleValTyrAlaMetValPheLysProHisAsnPheGluLeuGlyVal 875
Qy 1915 GAAGACCCCGCGCGCTTGTATATGAGGCGCGCGAGGCGCGAGGCGGTAATATC 1974
Dh 875 LysTyrProThrValLeuAsnValTyrGlyProGluValGlnThrValAsnAsnTh 895
Qy 1975 CTTCAAAGCGATCAAGTATGCGGCTCAACACACTGCGCTCCGCGCTACGCGCTGAT 2034
Dh 895 rPheLysGlyLysHisGlnLeuArgMetHisMetLeuAlaIleGlnGlyTyrCysValII 915
Qy 2035 TGTGATTGACGAGCGGCTCTGTACAGGAGGCTTCGGTTGCAAGGCGCGCTGAA 2094
Dh 915 eCysIleAspSerArgGlySerArgIleArgIleArgIleLysArgPheGlnSerHisIleArgI 935
Qy 2095 CCAATGGCGCGAGTGGAGTGAAGACGAGGCGCGCTGCGATGCTGCGCGGAA 2154
Dh 935 YArgMetGlyLeuValGluLeuThrAspGlnValAlaAspAlaLeuArgSerLeuSerAspGI 955
Qy 2155 GTATGGCTTCATCGACCTGAGCCGAGTGCATGCGCTGCTGCTGCGGCGCTTCT 2214
Dh 955 nLeuGlyLysIleAspMetLysArgValAlaIleHisGlyLysPheSerTyrGlyGlyTyrLe 975
Qy 2215 CTTCGCTATGGGCGCTAATCCACAAAGCCAGGTGTTCAAGTGGCGCTGCGGCGCTCC 2274
Dh 975 uSerLeuMetGlyLeuValGlnTyrProLysIlePheLysValAlaIleAlaIleGlyAlaPr 995
Qy 2275 GGATGCGCTGGATGGCGTGAACAGACAGGATACAGGCGCTGATGAGCGCTGCGCG 2334
Dh 995 oValThrAsnTrpGluTyrTyrAspThrGlyTyrThrGluArgTyrMetLysPheProGI 1015
Qy 2335 GAACAGCAGCAGCGTATGAGCGGCGGTTCGTGCGCTGACGCTGAGAGCTGCGCG 2394
Dh 1015 nAsnAsnGlnLysGlyTyrSerAlaGlySerValLeuGluTyrValAlaAsnSerPheProGI 1035
Qy 2395 TGAGCCCAACCGCTTATATCTCCACGCGCTTCGTGAGCAAAAGTGCATTTTCC 2454
Dh 1035 uGlnAspLysArgLeuLeuLeuIleHisGlyLeuIleAspGlnAsnValHisPheCysHI 1055
Qy 2455 CACAACCTCTCGTCCGCTCCCAAGTATCCGAGGAGGAGAACTTACAGCTCCAGATCTA 2514
Dh 1055 sThrSerArgLeuIleSerAlaLeuAsnLysAlaAsnLysProTyrGluValHisLeuPh 1075
```

```
Qy 2515 CCCAAGCAGAGACAGTATTCGCTGCCCGGAGTGGCGGCGAGCATGTGAATGACGTT 2574
Dh 1075 eProGluGlnArgHisSerLeuArgAsnLeuGlnSerAsnLysAsnTyrGluThrLysLe 1095
Qy 2575 GCTGCATTTTCTACAG 2590
Dh 1095 uLeuSerPheLeuGln 1100

RESULT 11
O9HBM3 PRELIMINARY: PRT: 465 AA.
AC O9HBM3;
DT 01-MAR-2001 (TRENBLREL. 16, Created)
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE Dipeptidyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221636; AAG29768.1; -.
DR MEROPS; S09.018; -.
FT NON_TER
SQ SEQUENCE 465 AA; 53197 MW; 22939BC0A4FE29CE CRC64;

Alignment Scores:
Pred. No.: 2,47e-91 Length: 465
Score: 1494.50 Matches: 281
Percent Similarity: 62.95% Conservative: 69
Best Local Similarity: 50.54% Mismatches: 105
Query Match: 31.17% Indels: 101
Gaps: 2
DB:

US-09-976-674-4 (1-2617) x O9HBM3 (1-465)
Qy 932 ACAGGACAGCAAGATCCCAAGATGCTTGAACATGCTGAGTTCACATGACAGCCAG 991
Dh 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
Qy 992 GCAAGATCGTCTGACCCAGCAGAGAGAGGCTGGTGCAGCGCTTCAGCTGCTGTCCG 1051
Dh 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheLeu 40
Qy 1052 AAGGTGAGTACATGCCAGGCGCGGTGACCCGCGATGCGCAATATGACCTGGCGCATG 1111
Dh 41 GlyValGluTyrIleAlaArgAlaGlyThrPheProGluGlyLysTyrAlaIlePheLe 60
Qy 1112 TTCTGGACCGCGCCACAGCATGCTGCAGCTGCTGCTCCCTCCCGCGCGCTGTATC 1171
Dh 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
Qy 1172 CCGAGCAGAGAAATGAGAGCAGCGGCTGACCTCTGCGAAGAGTGTCCCGAGAAATGC 1231
Dh 81 ProValGlnAspArgValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
Qy 1232 CAGCGTATGTGGGTGACGAGAGGCTCACAAGCTGTGATCAATGTTCATGACATCTTC 1291
Dh 101 ThrProLeuIleIleTyrGluGluThrThrAspIleThrIleAsnIleHisAspIlePhe 120
Qy 1292 TATCCCTTCCCCCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1351
Dh 121 HisValPheProGlnSerHis--GluGluGluIleGluPheIlePheIleAspIleSerGlnCys 139
```

```

OY 1352 AAGACCGGCTTCTGCAATTTGACAAAGTACACCGCGTTTAAATCCAGGCTACGAT 1411
    |||||
Db 140 LysThrGlyPheAlaGlnHisLeuTyrIleThrSerIleLeuLysGlnSerIleTyrLys 159
OY 1412 TGGAGTACGCGCTTACGCGCGGAGAGATGAATTAATGATGCGCCATTAAGAGAGAT 1471
    |||
Db 160 ArgSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGlnIle 179
OY 1472 GCTTCAGACCGGCGATGGAGATTTGGCGAGCGACGCGTCCAAAGATCTGGGTCAT 1531
    |||||
Db 180 AlaIleThrSerGlyGlyTyrPheIleuLysGlnArgHisGlySerAsnIleGlnValAsp 199
OY 1532 GAGGAGACAGCTGTACTTCCAGGCGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1591
    |||
Db 200 GluValAlaGlnLeuValTyrPheGlnGlyThrLysAspSerProLeuGlnHisIleu 219
OY 1592 TACGTGTACAGTATGAGCGCGCGCGAGAGATGACGCTCCACAGCGCGCGCTTCTCC 1651
    |||||
Db 220 TyrValAlaSerTyrValAsnProGlyGlnValThrArgLeuThrAspArgGlyTyrSer 239
OY 1652 CATAGCTCTCATAGCGCAGACATTCACATGTTCTGACGCGCTACAGACGCGTGAGC 1711
    |||||
Db 240 HisSerCysLysIleSerGlnHisCysAspPheIleSerLysTyrSerAsnGlnLys 259
OY 1712 AGCGCGCGCTGCGTACAGCTACAGCTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1771
    |||
Db 260 AsnProHisCysValSerLeuTyrLysLeuSerProGlnAspAspProThrCysLys 279
OY 1772 CAGCGCGCGCTTCTGCGTACAGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 1831
    |||||
Db 280 ThrLysGlnPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspGlyThrPro 299
OY 1832 CCAGGATCTTCATTTCCACACGCGCTCGGATGTCGCGCTACGCGCATGATCTACAG 1891
    |||||
Db 300 ProGlnIlePheSerPheGlnSerThrThrGlyPheThrLysGlyMetLeuTyrLys 319
OY 1892 CCCCAGCGCTTCTGCGCAGCGAGAGAGACCGCGCGCTCTTGTATATGAGCGCGCC 1951
    |||||
Db 320 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyPro 339
OY 1952 CAGGTGACGTGTGTAATTAATCTCTTCAAGGCGATCAAGTACTGCGCTCAACACACTG 2011
    |||
Db 340 Gln----- 340
OY 2012 GCCTTCCTGGGCTACGCGGTGTTGATGATGACGCGAGGCGCTCTGTCAGCGAGGCTT 2071
    -----
Db 340 ----- 340
OY 2072 CGGTCGAAGGGGCGCTGAAAAACAATGGCGCAGGTGGAGATGAGAGACCGGTGGAG 2131
    -----
Db 340 ----- 340
OY 2132 GGCGTCGATTCGTGCGCGAGAGATGATGCTTCATGACCTGAGCGAGTTCATTCAT 2191
    -----
Db 340 ----- 340
OY 2192 GCGTGGTCTTACGGGGGCTTCTCTCGCTCATGGGCTAATCCACAAGCCCGAGGTGTC 2251
    -----
Db 340 ----- 340
OY 2232 AAGGTGGCATCGGGGGGCGCGCGCGTCAACGCTGTGATGGCTACGACACAGGTACACT 2311
    |||||
Db 341 ---ValAlaIleAlaGlyAlaProValThrLeuThrPheThrLysAspThrGlyTyrThr 359
OY 2312 GAGCGCTACATGAGAGCTCCGTAGAACACACGACGCGCTATGAGCGGGGTTCCGTGGCC 2371
    |||||
Db 360 GluArgTyrMetGlnHisIleProAspGlnAsnGlnGlnGlyTyrTyrLeuGlnLysSerValAla 379
OY 2372 CTGACGCTGGAGAGAGCTGCGCATGAGCCCAACGCGTGTGCTTATCTCCACGGCTTCTG 2431
    |||||
Db 380 MetGlnAlaGlnLysPheProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeu 399

```

```

OY 2432 GACGAAAAAGTGCACACTTTTCCACAAAACCTCTGTCGTCCCAACTGATCGGACGAG 2491
    |||||
Db 400 AspGlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419
OY 2492 AAACCTTACCACTCCAGATTCACACCGGCGAGACACAGATTCGTCGCGCGAGTCG 2551
    |||||
Db 420 LysProTyrAspLeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGlnSer 439
OY 2552 GCGGAGCACTATGAGTACAGCTGCGTGCATTTCTACAGGAATTAAGTCTG 2599
    |||||
Db 440 GlyGlnHisTyrGlnLeuHisLeuLeuHisTyrLeuGlnGlnValAsnLeu 455

RESULT 12
O9NXF4 PRELIMINARY: PRT: 469 AA.
AC O9NXF4
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ20283 fis, clone HEP04088.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Oabayashi M.,
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000290; BAA91059.1; -.
DR MEROPS: S09.018; -.
SQ
SEQUENCE 469 AA: 54367 MW; 088ED0B1E46C1F CRC64;

Alignment Scores:
Pred. No.: 2e-83 Length: 469
Score: 1376.00 Matches: 245
Percent Similarity: 75.51% Conservative: 85
Best Local Similarity: 56.06% Mismatches: 105
Query Match: 28.70% Indels: 2
DB: 4 Gaps: 2

US-09-976-674-4 (1-2617) x O9NXF4 (1-469)
OY 80 TTCAGGTGACAGACACTGTTGGAGCGGCTCCGAGCATCATCCAGCGAGCGGCAAG 139
    |||
Db 35 PheTyrValGlnArgTyrSerThrSerGlnLeuLysLysLeuAlaAspThrArgLys 54
OY 140 TACTGGGCGCTATGTCAACAAGCGCGCGCGCGCGCGCTTGTGTCGAGAAGCGGAT 199
    |||
Db 55 TyrHisGlyTyrMetLeuAlaValProHisAspPheValAlaValAspArgAsnAsp 74
OY 200 GAGTGGGCGCGCGCGCGCGCGCGCTCTACTCTGCGGAATGCGCATATGCGAGCGGAGAG 259
    |||||
Db 75 ProAspGlyProHisSerAspArgIleTyrTyrLeuAlaMetSerGlyGlnAsnArgGln 94
OY 260 AACTCCCTCTCTACTCTGATGATCCCAAGAGTCCGGAAGAGAGGCTCTGCTCTCTG 319
    |||||
Db 95 AsnThrLeuPheTyrSerGlnIleProLysThrIleAsnArgAlaValLeuMetLeu 114
OY 320 TCCGTGAGCAATATGCTGATCATTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 379
    |||||
Db 115 SerTyrLysProLeuAsnProLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134
OY 380 GAGGAGAGCTGCTGAGGAGCGGAAACGCTGGGGGCTTGTGGCATGCTCTTACAGAC 439
    |||||
Db 135 GlnGlnGlnLeuLeuArgGlnArgLysArgIleGlyThrValGlyIleLeuLysSerTyrAsp 154
OY 440 TTCACAGCGAGAGTGGCGCTTCTCTTCCAGGCGCGAGACAGCGCTTCCAGCTGTGCG 499
    |||||
Db 155 TyrHisGlnGlnSerGlyThrPheLeuPheGlnAlaGlnLysSerGlyIleTyrHisValLys 174

```

QY 500 GAGGCGGCAAGAACGGCTTATGGTGTCCCTATGAACCGCTGGAAATCAAGACCCAG 559
 DB 175 AspIlglygProglnglyPheThrgInglProleuArpProAsnLeuValGluThrSer 194
 QY 560 TGGTCAGGCGCCCGGATGGACCCCAAAATCTGCCCTGGCCGCCCTCTTCCTCC 619
 DB 195 CysProAsnIleArgMetAspProLysLeuGlySerProAlaAspProAspTrpIleAlaPhe 214
 QY 620 ATCAATATACAGCAGCTTGGGTGGCCAAATCGACAGGCGAGGACCGCGCTGACC 679
 DB 215 ILeHISserAsnAspIleTrpIleSerAsnIleValThrArgGluGluArgLeuThr 234
 QY 680 TTCTGCCACCAAGTTATCCAAATGCTCGATGACCCCAAGTCTCGGGTGGCCACC 739
 DB 235 TyrValIHisAsnGluLeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThr 254
 QY 740 TTGCTATACAGGAAGTGGACCGCTTCACCTGGGTCTGGTGTGGCCCAACGCTCC 799
 DB 255 PheValIleuGluGluGluPheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGlu 274
 QY 800 TGGGAAGGTTTCAGAGGCGCTTCAGAGCGCTGCAATCTGTATGAGAGATGATGATCC 859
 DB 275 ThrThrProSerGlyGly---LysIleLeuArgIleLeuTyrGluGluAsnAspGluSer 293
 QY 860 GAGTGGAGTCAATTCACGTCCTCTCTGCTGCTAGAGAAGAGACGCGACTCGTAT 919
 DB 294 GluValIHisIleIleHisValThrSerProMetLeuGluThrArgArgAlaAspSerPhe 313
 QY 920 CGGTACCCCAAGACAGCAGACGAATCCCAAGATGCTTGAATGGCTGAGTTCAG 979
 DB 314 ArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMet 333
 QY 980 ACTGACCCCAAGGCAAGATGCTTCGACCCAGAGAAGAGAGTGTGTCACGCTTCAGC 1039
 DB 334 ILeAspAlaGluGluArgIleIleAspValIleAspLysGluLeuIleGlnProPheGlu 353
 QY 1040 TCCTGTTCCTCCGAAGGTGGATACATGCGCCAGGCGGGTGGACCCGGATGCAATAC 1099
 DB 354 ILeuLeuPheIleGluGluValGluTyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyr 373
 QY 1100 GCGTGGGCGATGTCCTGGACCGGCGCCAGAGTGTCTCGCTGCTGCTCCGCCG 1159
 DB 374 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGluIleValLeuIleSerPro 393
 QY 1160 GCGCTGTTCATCCGACACAGCAAGATGAGGACGCGCTAGCGCTGCCAGAGCTGTC 1219
 DB 394 GluLeuPheIleProValIleGluAspArgValIleMetGluArgGluArgLeuIleGluSerVal 413
 QY 1220 CCCAGGAATGTCCAGCGGTATGTGTGTACAGAGAGGTCACCAAGTGTGATCAATGT 1279
 DB 414 ProAspSerValThrProLeuIleIleTyrGluThrThrAspIleTrpIleAsnIle 433
 QY 1280 CATGACATCTTATCCCTCCCTCCCAATCAGAGGAGAGAGAGCTGCTTCTCCGCC 1339
 DB 434 HisAspIlePheHisValPheProGlnSerHis---GluGluGluIleGluPheIlePhe 452
 QY 1340 GCCAATGATGCAAGACCGGCTTCTGCCATTGTACAAAGTCAACGCGCGT 1390
 DB 453 AlaSerGluCysLysThrGlyPheArgHisLeuTyrLysIleThrSerIle 469
 RESULT 13
 Q9HBM2 PRELIMINARY; PRT; 360 AA.
 AC Q9HBM2;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Dipeptidyl peptidase 8 (Fragment).
 GN DPP8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NX NCBI_TaxID=9606.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20467194; Pubmed=11012666;
 RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
 RA Correll M.D.;
 RT "Cloning, expression and chromosomal localization of a novel human
 RT dipeptidyl peptidase (DPP) IV homolog, DPP8.";
 RL Eur. J. Biochem. 267:6140-6150(2000).
 DR EMBL: AF221637; AAC29769.1; -;
 DR MEROPS: S09.018; -;
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR Pfam: PF00326; Peptidase_S9; 1.
 FT NON_TER 1
 FT NON_TER 360
 SQ SEQUENCE 360 AA; 41070 MW; CF81C0BB61423E22 CRC64;
 Alignment Scores:
 Pred. No.: 4,21e-73 Length: 360
 Score: 1220.50 Matches: 236
 Percent Similarity: 54.31% Conservative: 54
 Best local Similarity: 44.19% Mismatches: 69
 Query Match: 25.45% Indels: 175
 DB: Gaps: 2
 US-09-976-674-4 (1-2617) x Q9HBM2 (1-360)
 QY 710 GATGACCCCAAGTCTGGCGGTGTGGCCCACTTCTGATACAGGAAGATTCGACCGCTTC 769
 DB 2 GluAspAlaArgSerAlaGlyValAlaThrPheValIleuGluGluPheAspArgTyr 21
 QY 770 ACTGGTACTGCTGGTGGCCACACGCTCTCGGGAAGTTCAGAGGCGCTCAACAGCTG 829
 DB 22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
 QY 830 CGAATCTGTATGAGAAAGTGTGATGAGTCCGAGGTGAGGTGATTCACGTCCCTCTCT 889
 DB 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
 QY 890 GCGGTAGAAGAAAGAGACGAGTCTGATGCGTATACCCAGACAGGCGCAAGATCC 949
 DB 61 MetLeuGluThrArgAlaArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 QY 950 AAGATTGCTTGAACCTGCTGAGTTCAGACTACAGACCCAGGCAAGATGCTTCGACC 1009
 DB 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluArgIleIle----- 98
 QY 1010 CAGAGAAAGAGCTGTGTCACACCTTCAGCTCGCTGTTCCGAAGTGGAGTACATGCC 1069
 DB 98 ----- 98
 QY 1070 AGGCGCGGTGACCCCGGATGCAAAATACGCTGGGCCATGTTCTGGACCGGCCAG 1129
 DB 98 ----- 98
 QY 1130 CAGTGGCTCAGCTGTCCTCTCCCGCGCCCTGTTCAATCCGAGACAGAGAAATGAG 1189
 DB 98 ----- 98
 QY 1190 GAGCAGCGGCTAGGCTCTGCGCAGAGCTGTCCCAAGAAATGCAACCCGATGTGTGTAC 1249
 DB 98 ----- 98
 QY 1250 GAGGAGTCAACCAACGCTGTGATCATGATCATCATCTTATCCCTTCCCCCAATCA 1309
 DB 98 ----- 98
 QY 1310 GAGGAGAGACGAGCTCTGCTTCTCCGCGCAATGAAATGCAAGACCGGCTTGTGCAT 1369
 DB 98 ----- 98
 QY 1370 TTGTACAAGTCAACCGCGCTTTTAAATCCAGGCGTACATTTGGAGTGAAGCTTCAGC 1429

```

Db 98 ----- 98
OY 1430 CCGGGGAAGATGATTTAAGTCCCATTAAGAAAGATTTGCTCTGACAGCGGTGA 1489
Db 98 ----- 98
OY 1490 TGGAGATTGGCGAGCGACGGCTCCAAAGATCTGGTCAATGAGAGACCAAGCTGTG 1549
Db 99 -----ValasgIValaIArgatgIeuVal 106
OY 1550 TACTTCAGAGGACCAAGAGACAGCGCGTGGAGACACACCTTACGTGGTCAATAG 1609
Db 107 TTTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669
OY 1610 GCGGCGGCGAGAGTGTACAGCTCAGCAGCGCGCGCTTCCCATAGCTGCTCCATGAC 1669
Db 127 AsnProglgIValThlrArgleuthrAspArgIyTyrSerHisSerCysCysIleSer 146
OY 1670 CAGAACTTCGACATGTTGCTGTCGACATACAGCGGTGAGCAGCGCGCGCTGCGAC 1729
Db 147 GlnHisCysAspPhePheIleSerIyTyrSerAsnGlnIyAsnProHisCysValSer 166
OY 1730 GTCTACAAAGTGAAGGCGCGCGACAGCAGCGCGCTGACACAGCGCGCTTGTGGCT 1789
Db 167 LeuTyrIyIleuSerSerProgluAspAspProthrcCysIySerThrlYsgIuPheTrpAla 186
OY 1790 AGCATGATGGAGGACAGCGCGCGCGATTTATGTTCTCCAGAGATCTTCATTTC 1849
Db 187 ThrIleuAspSerAlaGlyProleuProAspTyrThrProProgluIlePheSerPhe 206
OY 1850 CACAGCGCGTGGATGGCGCGCTACGCGATGATTCACAGCGCGCGCTTGGACCA 1909
Db 207 GluSerThrThrGlyPheThrIleuTyrGlyMetIleuTyrIyLysProHisPleuGlnPro 226
OY 1910 GGGAGAAGACACCGCGCGCTCTTTATATGAGAGCGCGCGCGCTGAGT 1969
Db 227 GlyIyLysTyrProthValIleuPheIleTyrGlyIyProgluIValGlnIleuValAsn 246
OY 1970 AACTCTTCAAAGGATCATAGTACTGGCGCTCAGACACACCTGGCTCCCTGGGCTACGCC 2029
Db 247 AsnaIyAspPheIyGlnIyValIyLysTyrPheArGlyeuaSnThrlleuAlaSerIeuIyTyrVal 266
OY 2030 GTGTTGATGATGAGCGCGCGCGCTCTGTCAGCGAGCGCTCGGTTGCAAGGGCGCTCG 2089
Db 267 ValIyValIleAspAsnArgIySerCysHisArgGlyeuaSnThrlleuIyIleuAlaPhe 286
OY 2090 AAAAAACAATGGCGCGAGGTGAGATCGAGACAGCAGGTGAGAGCGCTGCAATTCGTGCC 2149
Db 287 LysTyrIyLysMetGlnIyGlnIleGlnIleAspAspGlnIyValGlnIyIleuIyTrpIleuAla 306
OY 2150 GAGAAGTATGGCTTATGACACTGAGCGGAGTGGCCATTCAGTGGCTGCTACGGGGGCT 2209
Db 307 SerArgTyrAspPheIleAspPheAspArgValIyGlnIleHisGlnIyTrpSerTyrGlyGly 326
OY 2210 TTCCCTGCGCTGATGGGCTATATCCACAAGCGCGCGCTGTCAGGTGCGCAATCGGGGCT 2269
Db 327 TyrIleuSerIeuMetAlaIleuMetGlnArgSerAspIlePheArgValAlaIleuAlaGly 346
OY 2270 GCCCGGTCACCGCTGTGATGGCTACGACACAGGGGTACACT 2311
Db 347 AlaProValThrIleuTrpIlePheTyrAspThrIyTyrThr 360
RESULT 14
O9HBM4
ID O9HBM4 PRELIMINARY: PRT: 310 AA.
AC O9HBM4:
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Dipeptidyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollett E., Sutherland G.R., McCaughan G.W.,
RA Correll M.D.,
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL: AF221635; AAC29767.1; -.
DR MEMOPS: S09.018; -.
DR InterPro: IPR001375; Peptidase_S9.
DR Pfam: PF00326; Peptidase_S9; 1.
FT NON_TER 1
SQ SEQUENCE 310 AA: 35396 MW; 8E87C34026D9C7AC CRC64;

Alignment Scores:
Pred. No.: 1,17e-64 Length: 310
Score: 1093.50 Matches: 200
Percent Similarity: 69.34% Conservative: 42
Best Local Similarity: 57.31% Mismatches: 58
Query Match: 22.81% Indels: 49
DB: Gaps: 1

US-09-976-674-4 (1-2617) x O9HBM4 (1-310)
OY 1553 TTCSAGGGGACCAAGAGACAGCGCGTGGAGACACACCTTACGTGGTCAATAGAGCG 1612
Db 1 PheGlnGlyThrIyAspSerProleuGlnHisIleuTyrValIyValSerTyrValAsn 20
OY 1613 GCGGCGAGATGTCAGCTGCTCAGCAGCGCGCGCTTCCCATAGCTGCTCAGAGCGAG 1672
Db 21 ProgluIyValThlrArgleuthrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
OY 1673 AACTTCGACATGTTGCTGTCGACATACAGCGGTGAGCAGCGCGCGCTGCGACAGC 1732
Db 41 HisCysAspPhePheIleSerIyTyrSerAsnGlnIyAsnProHisCysValSerIeu 60
OY 1733 TACAGCTGAGGCGCGCGCGCGACAGCAGCGCGCTGACACAGCGCGCGCTTGGGCTAGC 1792
Db 61 TyrIyIleuSerSerProgluAspAspProthrcCysIyThrlYsgIuPheTrpAlaThr 80
OY 1793 ATGATGAGGACAGCAGCTGCCCGCGGATTTATGTTCTTCCAGAGATCTTCATTTCAC 1852
Db 81 IleuAspSerAlaGlyProleuProAspTyrThrProProgluIlePheSerPheGln 100
OY 1853 ACGCGCTGCGATGCGCGCTTACAGCGATGATTCATTAAGCGCGCGCTTGGAGCGAG 1912
Db 101 SerThrThrGlyPheThrIleuTyrGlyMetIleuTyrIyLysProHisPleuGlnProGly 120
OY 1913 AAGAAGACACCGCGCTCTTTATATGAGAGCGCGCGCTGAGCTGGATTAAC 1972
Db 121 LysIyTyrProthValIleuPheIleTyrGlyIyProgluIn----- 134
OY 1973 TCCCTTCAAGGATCAAGTACTTGGCGCTCAACACAGCTGCCCTCCCTGGGCTACGCCGTG 2032
Db 134 ----- 134
OY 2033 GTTGTGATTGAGCGGCGCGCGCTCTGTCAGCGAGCGCTCGGTTGCAAGGGCGCTGAAA 2092
Db 134 ----- 134
OY 2093 AACCAATGGCGCGAGTGGAGATCGAGACACAGGTGAGAGCGCTGCAATTCGTGGCGAG 2152
Db 135 -----GlyGlnIleIleuIleAspAspGlnIyValGlnIyIleuIyTrpIleuAlaSer 151
OY 2153 AAGTATGCTTATGACACTGAGCGGAGTGGCCATTCAGTGGCTGCTACAGGGGCGCTTC 2212
Db 152 ArgTyrAspPheIleAspPheAspArgValIyGlnIleHisGlnIyTrpSerTyrGlyGlyTyr 171
OY 2213 CTTCTGCTCATGAGGCTATATCCACAAGCGCGCGTTCGAAGGTGCGCATCGCGGCTGCC 2272

```


Db 387 GluThrIleThrArgAlaGlyPhePheSerAspGlyThrValThrValGlnValMet 406
 QY 1118 GACGGCCCGCCAGAGTGGCTCCAGCTGCTCCCTCCCGCCCGCCCGCTTCATCCCGAGC 1177
 Db 407 SerArgAspGlnAlaGlnCysSerLeuLeuLeuIleProTyrThrAspPheLeuPro 426
 QY 1178 ACAGAG----- 1183
 Db 427 GluGluLeuGlyGlySerIleLysGlnAspAsnLeuGlnLeuSerThrAspLeuAsnMet 446
 QY 1184 -----AATGAGAGAGCAGCGGTAGCTCTGCGCAGAGCTGCCAGC----- 1225
 Db 447 GlyValThrAspAspLysSerHisGluGluThrMetGluLysProProArgGlyLysLeu 466
 QY 1226 -----AATGTCACGCCGTATGTGGTGTAGCAGAGGTCACAGACGTCGATCATATTT 1279
 Db 467 ArgGlyThrValGln-----IleHisLysAlaArgAsnAspTyrTrpIleAsnThr 483
 QY 1280 CATGACATCTTCTATCCCTTCCCGCCCATGAGAGGAGAGCAGAGCTGCTGC---TTCTCTC 1336
 Db 484 HisAsnAlaIleTyrProLeuLysIleThrAspGluGlnHisProMetCylGluPheIle 503
 QY 1337 CGCGCCAAATGATGCAGACAGCGGCTTGTCCATTTGTACAAAGTCACCGCGTTTAAAA 1396
 Db 504 TyrCysLeuGluLysProAsnGlySerCys---LeuAlaLeuIleSerAlaGluLeuAsp 522
 QY 1397 TCCACAGGCTTCGATGGAGTGGAGTGCAGGCCCTTCAGCCCGGGAAGATTTAAGTCCCC 1456
 Db 523 GlnAsnGlyTyr-----CysArg 528
 QY 1457 AATTAAGCAATGATGTCTGTGACAGCGGTGAATGGAGGTTTGGCAGCGCGCTCC 1516
 Db 529 HisThrGluGluLysLeuLeuMetAlaGlnAsnPheSerIleAsnLysSerMetCyl--- 547
 QY 1517 AAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCACCAGAGACAGCGCG 1576
 Db 548 ---IleValValAspGluValArgGluLeuValTyrTyrValAlaAsnGluSerHisPro 566
 QY 1577 CTGAGACACCACTCTACGTGTGACGTATGAGGCGCGCGCGAGATGTACGCTCAGC 1636
 Db 567 ThrGluThrAsnIle---CysValSerHisTyrArgThrGlyGlnHisAlaGluLeuThr 585
 QY 1637 ACGCCCGGCTTCTCCCATAGCTGTCTCCATGAGCCAGAAC-----TTC 1678
 Db 586 GluSerGlyIle-----CysPheLysSerGluArgAlaAsnGlyLysLeuAlaLeu 602
 QY 1679 GACATG-----TTGTCGACCCACTACAGCAGCGGTAGCAGCGCGCTGCGTGCAC 1729
 Db 603 AspLeuAspHisGlyPheAlaCysTyrMetThrSerValGlySerProAlaGluCysArg 622
 QY 1730 GTTACAAAGCTGAGCGCGCGCCGAGCAGACCCCTGACAAAGCAGCGCGCTTGGGCT 1789
 Db 623 PheTyrSerPheArgTyrLysGlnAsnGluValLeu-----ProSerThrValTyrAla 640
 QY 1790 AGCATGATGAGGAGCAGCGCTGCC-----CCGAT-----TATGTTCTCCAGAG 1837
 Db 641 AlaAsnIleThrValSerGlyHisProGlyGlnProAspLeuHisPheAspSerProGlu 660
 QY 1838 ATCTTCATTTCCACAGC---CGCTCGATGTGGCGCTACGCGCATGTCTACAAAGCC 1894
 Db 661 MetIleGluPheGlnSerLysThrGlyLeuMetHisTyrAlaMetIleLeuArgPro 680
 QY 1895 CACGCGTTGACAGCGGGAAGAAGACACCCCGCTCTTGTATGAGAGGCCCCCAG 1954
 Db 681 SerAsnPheAspProTyrLysLysTyrProValPheHisTyrValTyrGlyLysProGly 700
 QY 1955 GTGACGCTGTGAATTAATCTCTCAAGCATCACTAGTACTGGCGCTCAACACACTGGCC 2014
 Db 701 IleGlnIleValHisAsnAspPheSerTrpIleGlnTyrIleArg-----PheCys 717
 QY 2015 TCCCTGGGCTACGCGGTGTGTGATTGAGCGAGCGGCTCTGACGAGAGGCTTGG 2074
 Db 718 ArgLeuGlyTyrValValValPheIleAspAsnArgLysSerAlaHisArgGlyIleGlu 737

QY 2075 TTGCAAGGCGCCCTGAAAAACCAATGGGCCAGTGGAGATCGAGACAGGTGGAGGCG 2134
 Db 738 PheGluArgHisIleHisLysLysMetCylThrValGlnValGlnAspGlnValGluGly 757
 QY 2135 CTGCAGTTCGTGGCCGAGAGATAT---GGCTTCATGACACTGAGCGCGAGTTGCCATTCAT 2191
 Db 758 LeuGlnMetLeuAlaGlnArgThrGlyGlyPheMetAspMetSerArgValValHis 777
 QY 2192 GCGTGGCTTACGGGGGCTCTCTGCTCATGGGCTAATCCAAAGCCCGAGGTGTC 2251
 Db 778 GlyTyrSerTyrGlyGlyTyrMetAlaLeuGlnMetIleAlaLysHisProAsnIleTyr 797
 QY 2252 AAGTGGCCATCGGGGCGGCGCCCGTCAACCGCTGAGAGGCTAGAGACAGGTAACCT 2311
 Db 798 ArgAlaAlaIleAlaGlyGlyAlaValSerAspTrpArgLeuTyrAspThrAlaTyrThr 817
 QY 2312 GAGCGCTACATGAGACGTCCCTGAGAAACACAGCAGCGCTATGAGCGGGTTCGCGGCC 2371
 Db 818 GluArgTyrMetGlyTyrPro---LeuGluGlnHisValTyrGlyAlaSerSerIleThr 836
 QY 2372 CTGCAGTGGAGAGCTGCCCATGAGCCCAACCGCTTATCTCCACGCTTCTCTG 2431
 Db 837 GlyLeuValGluLysLeuProAspGluProAsnArgLeuMetLeuValHisGlyLeuMet 856
 QY 2432 GACGAAACGTGCACTTTTTCACACAAACTCTCTGCTCCCACTGATCCGAGCAGG 2491
 Db 857 AspGlnAsnValHisPheAlaHisLeuThrHisLeuValAspGluCysIleLysLysGly 876
 QY 2492 AAACCTTACAGCTCTCAGATCTACCCCAAGCAGAGACACAGATATGCTGCCCGAGTGC 2551
 Db 877 LysTrpHisGluLeuValIlePheProAsnGlnAlaGlnHisGlyValArgAsnAsnAspAla 896
 QY 2552 GCGCAGCACTAAGATCACTGCTGCACTTTCTACAGGAA 2593
 Db 897 SerIleTyrLeuAspAlaArgMetMetTyrPheAlaGlnGln 910

Search completed: December 12, 2002, 11:51:46
 Job time : 108 secs

Db 379 ProhiscysvalSerleuTyrlsLeuSerProgluAspProthrcysLysThr 398
 QY 1775 CCCCCTTCCTGGGTACATGATGAGGAGCCAGCTGCCCCCCCGATATATCTCTCA 1834
 Db 399 LysgluTherPalatThrIleLeuAspSerAlaGlyProLeuProAspPyrThrPro 418
 QY 1835 GAGATCTTCATTTCCACAGCCGCTGATGTGCGGCTTACGAGCATACAGACC 1894
 Db 419 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrlGlyMetLeuTyrlsPro 438
 QY 1895 CAGCCCTTCAGCCAGGAGGAGAGACACCCAGCTCTCTTGTATATGAGGCCCCAG 1954
 Db 439 HisAspLeuInProGlyLysTyrlsTyrlProthValLeuPheIleTyrlGlyProGln 458
 QY 1955 GTGACGTGTGATAATCTCTTCAAGAGCATCACTACTGCGGCTCAACACTGTGACC 2014
 Db 459 ValGlnLeuValAsnAsnArgPheLysGlyValTyrlsTyrlPheArgLeuAsnThrLeuAla 478
 QY 2015 TCCCTGGGCTACGCGCTGTGTGTGATTGACGCGAGGGCTCTGTACAGGAGGCTTCGG 2074
 Db 479 SerLeuGlyTyrlValValValIleAspAsnArgLysSerCysHisArgLysLeuLys 498
 QY 2075 TTGCAAGGGCCCTGAAACCAATGGCGCCAGCTGAGATGAGAGACAGGTGAGGGC 2134
 Db 499 PheGlnGlyAlaPheLysTyrlsMet----- 507
 QY 2135 CTGCACTTCGTGGCCGAGAGTATGCTTCATCGACTGAGCCGAGTTCATCCATGAGC 2194
 Db 507 ----- 507
 QY 2195 TGGTCTACAGGGGCTTCTCTGCTCAGTGGGCTATCCACAGCCCGAGTGTTCAG 2254
 Db 507 ----- 507
 QY 2255 GTGCGCATCGGGGTGCCCCGCTCACGCTGTGATGAGCTACAGACAGAGTACAGTGA 2314
 Db 508 ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrlAspThrGlyTyrlGln 527
 QY 2315 CGCTACATGAGACCTGCTGAGAACACACGAGCTATGAGGGGCTTCCTGAGCCCTG 2374
 Db 528 ArgTyrlMetGlnIleAspProAspGlnAsnGlnGlyTyrlsTyrlLeuGlySerValAlaMet 547
 QY 2375 CACGTGAGAGAGCTGCCAATGAGCCCAACGCTGTATCTCCAGGCTTCCTGGAGC 2434
 Db 548 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 567
 QY 2435 GAAAGCTGCACTTTTCCACACAACTCTGCTCCCACTGATCCGAGACAGGAGAA 2494
 Db 568 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 587
 QY 2495 CCTTACAGCTCCAGCTCTCCCAACGAGACAGATATGCTGCTGCCCCAGTCCGGGC 2554
 Db 588 ProTyrlAspLeuGlnIleTyrlProGlnGluArgHisSerIleArgValProGlnSerGly 607
 QY 2555 GAGCACTATGAGTCACTGCTGCACTTTCTACAGAAATACCTC 2599
 Db 608 GlnHisTyrlGlnLeuHisLeuHisTyrlLeuGlnGlnAlaSerLeu 622

RESULT 7
 ID 075868 PRELIMINARY; PRT: 432 AA.
 AC 075868;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE R33083_1 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,

RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J.,
 RA Dangnan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
 RA Atlix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
 RA Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RA Olsen A.S., Garrino A.V.,
 RT "Sequence analysis of a 2.5 kb region in 19p13.3."
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005783; AAC62840.1;
 FT NON_TER 432
 SO SEQUENCE 432 AA; 48595 MW; 64E2B85BE0523A7E CRC64;

Alignment Scores:
 Pred. No.: 1,07e-115 Length: 432
 Score: -1859.50 Matches: 354
 Percent Similarity: 97.52% Conservative: 0
 Best Local Similarity: 97.52% Mismatches: 0
 Query Match: 38,78% Indels: 9
 DB: 4 Gaps: 2

US-09-976-674-4 (1-2617) x 075868 (1-432)

QY 11 ATGCCACACCGGGACCCCAAGCCGACGAGGCGAGCGCCGACAGATGACCG 70
 Db 72 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaAlaThrAspAspPro 91
 QY 71 GCGCGCGCTCCAGGTGAGAGACACGTCGTGGAGCGGCTCCGAGCATATCCAGCGC 130
 Db 92 AlaAlaArgPheGlnValAlaGlnHisSerThrPaspGlyLeuArgSerIleLeuHisGly 111
 QY 131 AGCCGCAAGTACTCGGGCTCATGTGTCAACAAGGCGCCCACTTCCAGTTGTGTCAG 190
 Db 112 SerArgLysTyrlSerGlyLeuLeuLeuValAsnLysAlaProHisAspPheGlnValGln 131
 QY 191 AAACGAGATGATGTGGGCCCCATCCACGCGCTCTACTACCTGGAGATCCATATGAGC 250
 Db 132 LysThrAspGlnSerGlyProHisSerHisArgLeuTyrlsTyrlGlnGlyMetProTyrlGly 151
 QY 251 AGCCGAGAGAACCTCCCTCTACTCTGATGATTCGCAAGAAGTCCGGAAGAGAGCTCTG 310
 Db 152 SerArgGlnAsnSerLeuLeuTyrlSerGlnLeuProLysValArgLysGlnAlaLeu 171
 QY 311 CTGCTCTCTCTCGAGACAGATGCTGATCTTCACAGCCAGCCGACATGGGCTC 370
 Db 172 LeuLeuLeuSerThrLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 191
 QY 371 TACTCTCGGAGAGAGAGCTGCTGAGGAGCGGAAACGCTGGGGGCTTCGGGATCACC 430
 Db 192 TyrSerArgGlnGlnGlnLeuLeuArgGlnArgLysArgLeuGlyValPheGlyIleThr 211
 QY 431 TCTACGACTTCCACAGCAGAGATGGCTTCTCTTCCAGGCGACAGACAGCTCTTC 490
 Db 212 SerTyrlAspPheHisSerLysSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 231
 QY 491 CACTGTGCGGAGCGGCGGCAAGACGGCTCAG-----GTCTCC 529
 Db 232 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProGlyProGlyCysValSer 251
 QY 530 CCTATGAACCGCTGGAATCAAGACCAAGTCTCGAGGCGCCCGATGAGCCCAATATC 589
 Db 252 ProMetLysProLeuGlnIleLysThrGlnCysSerGlyProArgMetAspProLysIle 271
 QY 590 TGGCTTCGCGAGCCCTCTCTTCTTCTTCAATCAACAGCAGCTGTGGGTGGCCAC 649
 Db 272 CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuThrValAlaAsn 291
 QY 650 ATCGAGACAGGAGAGAGCGGCGGTGACCTTCTCCACAGGTTATCCAAATGCTCTG 709
 Db 292 IleGlnThrGlyGlnGlnArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 311
 QY 710 GATGACCCCAAGTCTGGGGGTGGCCACTTGTCTATACAGAGAGAGTGTGACCGCTTC 769

Db 312 AspasProlyseralaglyValAlaThrPheValIleGlnGluIupheaspargPhe 331
 QY 770 ACTGGTACTGTGGTGGCCACAGCCTCTGGAGGTTCAGAGGCGCTCAAGCGTG 829
 Db 332 ThrGlyYrtrPrpCysProThrAlaSerTrpGlu-----GluIuLeuYsrThrLeu 349
 QY 830 CGAATCCTGATAGAGAGTGCATGATGATCCGAGGTGAGGTCAATTCACCTCCCTCTCT 889
 Db 350 ArgIleuYrGlnGluValAspGluSerGluValGluValIleHisValProSerPro 369
 QY 890 GCGGTAGAGAGAGAGAGAGAGAGTCTGATACGCTACCCAGAGAGAGAGAGAGATCC 949
 Db 370 AlaLeuGluIupArgYsrThrAspSerTrpArgYrProArgThrGlySerIysAsnPro 389
 QY 950 AAGATTCCTTGAACAGTGGCTGATTCAGACTGACAGCAGGAGAGAGAGAGAGAGTCTGAC 1009
 Db 390 LysIleAlaLeuYsrGlnGluValGlnPheGlnThrAspSerGlnGlySerIleValSerThr 409
 QY 1010 CAG 1069
 Db 410 GlnGluYsrGlnGluValGlnProPheSerSerLeuPheProIysValGluTrpIleAla 429
 QY 1070 AGGGCCGGG 1078
 Db 430 ArgAlaGly 432

RESULT 8

Q96NT8 PRELIMINARY; PRT; 312 AA.

AC Q96NT8;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE CNA FLJ30094 f1s, clone BNGH4100034, weakly similar to dipeptidyl
 DE peptidase IV (EC 3.4.14.5).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxID=9606;
 RX 1)
 RP SEQUENCE FROM N.A.
 RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
 RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
 RA Kikuchi H., Kanda K., Nagatsuma M., Murakawa K., Kanehori K.,
 RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
 RA Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.;
 RT "MEDO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK054656; BAB70784.1;
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR000379; Ser ester_s site.
 DR Pfam: PF00326; Peptidase_S9; 1
 SQ SEQUENCE 312 AA: 35518 MW; ABB940AFC5877717 CRC64;

Alignment Scores:

Pred. No.: 1,88e-103
 Score: 1675.00
 Percent Similarity: 99.684
 Best Local Similarity: 99.684
 Query Match: 34.938

US-09-976-674-4 (1-2617) x Q96NT8 (1-312)

QY 1664 ATGAGCAGAACTTCGACATGTCTGACAGCACTACAGAGCGGTGAGAGAGCGCCGCTTC 1723
 Db 1 MetSerGlnAsnRheAspMetRheValSerHisTrpSerSerValSerThrProProCys 20
 QY 1724 GTGACAGCTTACAGAGTGAAGCGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1783
 Db 21 ValHisValIuYrGlyLeuSerGlyProAspAspProLeuHisIuYsrGlnProArgPhe 40

QY 1784 TGGGCTAGCATGATGAGAGCAG 1843
 Db 41 TrpAlaSerMetIleGluAlaIleAspCysProProAspYrValProProGluIlePhe 60
 QY 1844 CATTTCCACAGCGGCTCGAGATGGGGCTTACGGGAGATGATCAACAGCCGAGCGCTTG 1903
 Db 61 HisPheHisThrArgSerAspValArgLeuYrGlyMetIleYrLysProHisAlaLeu 80
 QY 1904 CAGCCAGGAG 1963
 Db 81 GlnProGlyLysIysHisProThrValLeuPheValYrGlyGlnProGlnValGlnLeu 100
 QY 1964 GTGATATACCTCTCAAG 2023
 Db 101 ValAsnAsnSerPheLeuSerIleLeuYsrIleuArgLeuAsnThrLeuAlaSerLeuGly 120
 QY 2024 TACGCGGTGGTGTGATATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2083
 Db 121 TyrAlaValValIleAlaSerGlyArgGlySerCysGlnArgGlyLeuArgPheGlnGly 140
 QY 2084 GCGCTGAAAGAACAAATGGGCGCAGGTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2143
 Db 141 AlaLeuLysAsnGlnMetIleGlnValGluIleGlnAspGlnValGlnGlyLeuGlnPhe 160
 QY 2144 GTGGCCGAGAGATAGGCTTCAATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2203
 Db 161 ValAlaGlnLysTrpGlyLeuIleAspLeuSerArgValAlaIleHisGlyTrpSerTrp 180
 QY 2204 GGGGCGTCTCTCTCGCTCATGGGCTTAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2263
 Db 181 GlyGlyPheLeuSerLeuMetIleLeuIleHisLysProGlnValPheLysValAlaIle 200
 QY 2264 GCGGAGGCGCCGATACAGAGTGTGATGGGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2323
 Db 201 AlaCysAlaProValThrValIleTrpMetAlaTrpAspTrpGlyTrpGlnAlaTrpYrMet 220
 QY 2324 GAGCTCCCTGAGAACAG 2383
 Db 221 AspValArgGluAsnAsnGlnHisGlyTrpGlnAlaGlySerValAlaLeuHisValGlu 240
 QY 2384 AAGCTCCCAATGAG 2443
 Db 241 LysLeuProAsnGlnIleProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnVal 260
 QY 2444 CACTTTTCCACAGAACTTCTCTGCTCCCACTGATCCGAGAGAGAGAGAGAGAGAGAGAGAG 2503
 Db 261 HisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGln 280
 QY 2504 CTCACATCTACCCAG 2563
 Db 281 LeuGlnIleTrpArgAsnGlnArgHisSerIleArgCysProGlnSerGlyGlnHisTrp 300
 QY 2564 GAAGTCAGCTGCTGACATTTCTTACAGAGATACCTG 2599
 Db 301 GluValThrLeuLeuHisPheLeuGlnGluTrpLeu 312

RESULT 9

Q9VC20 PRELIMINARY; PRT; 1042 AA.

AC Q9VC20;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE CG3744 protein.
 GN CG3744
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN 1)
 RP SEQUENCE FROM N.A.


```

Db 653 TygLylylProGlnValGlnLeuValAsnAsnArgPheLyGlyValIlystYrPheArg 672
OY 2000 CTCACACACTGGCCCTCCCTGGGCTACGCCCTGGTGTGATTTAGCGGCGGCTCTGT 2059
Db 673 LeuAsnThrLeuAlaSerLeuGlyTyrValIValIValIleAspAsnArgIlySerCys 692
OY 2060 CAGGAGGCTGGCTTCGAGGAGGCGCTGAAAACCAATAGGCGAGTGTGAGATTCAG 2119
Db 693 HisArgGlyLeuLeuysPheGlnGlyAlaPheLyTyrLysMetGlyGlnIleGlyLeuAsp 712
OY 2120 GACCAAGTGGAGGCGCTGCATTCCTGCGCGGAGAACTAGCTTCATGACCTGAGCCGA 2179
Db 713 AspIleValIleGlnGlyLeuGlnIlyLeuAlaSerArgTyrAspPheIleAspArgIly 732
OY 2180 GTTCCCATTCATGGCTGGCTCTACGGGCGCTCTCTCTGCTATGAGGCTAATCCACAAG 2239
Db 733 ValGlyLeuHisIleGlyTyrPheSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752
OY 2240 CCCCAGCTGTTCAGAGTGCATTCGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2299
Db 753 SerAspIlePheArgValAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAla 772
OY 2300 ACAGGCTACCTAGAGGCTACATGAGAGCTCCCTGAGACACACAGCGGCTATGAGGCG 2359
Db 773 ThrIlyTyrThrGlnIlyArgTyrMetGlyLysIleAspGlnIleGlnIlyTyrTyrLeu 792
OY 2360 GGTTCCGTGGCGGCGGCGGCGGAGAGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2419
Db 793 GlySerValIleAlaMetGlnIleGlnIlyLysPheProSerGlnProAsnThrLeuLeuLeu 812
OY 2420 CAGGCTCTCTGAGCAAGAAAGCTGCACTTTTTCACAAACATTCCTGCTCTCCCAATG 2479
Db 813 HisIleGlyLeuLeuysPheGlnValIleHisPheAlaIleHisThrSerLeuLeuSerPheLeu 832
OY 2480 ATCCGACACAGGAACTTCACAGCTCCAGATTCACCCCAAGAGACAGACAGATTCGCG 2539
Db 833 ValArgAlaIlyLysProIlyLysPheLeuGlnIleTyrProGlnIlyArgIleHisThrLeu 852
OY 2540 TGCCCGAGTGGCGGCGGAGCACTATGAGTCAAGTCAAGTCTGCTCTCTACAGAAATACCTG 2599
Db 853 ValProGlnSerGlyGlnIlyIlyTyrGlnLeuHisLeuLeuHisIlyTyrLeuGlnIlyLysLeu 872

RESULT 4
OY 075273 PRELIMINARY; PRT; 508 AA.
AC 075273;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE R26984_1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 11]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schulze K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garmes J.,
RA Danghan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Altix C., Andreise T., Frankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Kromayashi B., Arellano A., Montgomerie M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.,
*Sequence analysis of a 2.5 kb region in 19p13.3.
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC005594; AAC33801.1;
DR MEROPS; S09.019;
DR InterPro; IPR002410; Pro-annoptase.
DR InterPro; IPR000379; Ser-estrs_site.
DR PRINTS; PRO0793; PROANNOPTASE.
FT NON_TER 1 1

```

```

SO SEQUENCE 508 AA; 57750 MW; 2F4D6645BE2D2C89 CRC64;
Alignment Scores:
Pred. No.: 3,36e-163 Length: 508
Score: 2572.00 Matches: 493
Percent Similarity: 91.34% Conservative: 3
Best Local Similarity: 90.79% Mismatches: 3
Query Match: 53.64% Indels: 44
DB: 4 Gaps: 5

US-09-976-674-4 (1-2617) x 075273 (1-508)
OY 998 ATGCTGTCACCCAGAGAGAGAGAGTGTGAGCGCTTCAGCTGCTGTCGCCGAGAGCTG 1057
Db 1 IleValSerThrGlnGlnIlyGlnIlyLeuValGlnProPheSerSerLeuProIlyVal 20
OY 1058 GAGTACATGCCAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1117
Db 21 GlnTyrIleAlaIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIly 33
OY 1118 GACGCGCGCGCGAGAGTGGCTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1177
Db 34 AspArgProGlnGlnIlyPheLeuGlnIlyLeuValIleLeuProPheAlaLeuPheIleProSer 53
OY 1178 ACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
Db 54 ThrGlnAsnGlnGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIly 73
OY 1238 TATGCTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297
Db 74 TyrValValIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIly 93
OY 1298 TTTCCCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
Db 94 PheProGlnIleSerGlnIlyGlnIlyAspIlyLeuIlyLeuIlyLeuIlyLeuIlyLeuIlyLeuIly 113
OY 1358 GCGCTTGCATTTGTACAAAGTCAAGCGCGCTTTAAATCCAGGCTACAGATTGAGAT 1417
Db 114 GlyPheCysHisLeuIlyTyrValIlyThrAlaValIlyLeuIlySerGlnIlyTyrAspTrpSer 133
OY 1418 GACGCTTACGCGCGGAGAGATGAATTAAGTCCCAATTAAGAAAGATTCAGCTG 1477
Db 134 GluProPheSerProGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIlyGlnIly 145
OY 1478 ACCAGCGTGAATGGAGAGTGTGGCGAGGAGCGGCTCCAGATCTGCTGATGAGAG 1537
Db 146 ThrAsnAla-----IleTyrValIleGlnIly 154
OY 1538 ACCAAGCTGTACTTCCAGGAGCAAGAGACAGCGCGCTGAGACACACCTCTACGCTG 1597
Db 155 ThrIlyLeuValTyrPheGlnIlyThrIlyAspTrpProLeuGlnIlyHisIleTyrVal 174
OY 1598 GTACAGTATGAGCGCGCGCGGAGATCTACGCGCTACAGCGCGCGGCTCTCCATAGC 1657
Db 175 ValSerTyrGlnAlaIleGlnIlyGlnIlyValIlyArgLeuThrThrProGlnIlyPheSerHisSer 194
OY 1658 TGCTCCATAGACCCAGAACTTCGACATGTTCTGACAGCACTACAGACCGTGTAGACGCGG 1717
Db 195 CysSerMetSerGlnAspPheAspMetPheValSerHisTyrSerValIleThrPro 214
OY 1718 CCTGCGTGCAGCTGTACAGAGTGTAGCGCGCGCGAGCGAGCGAGCGCGCGCGCGCGCGCGCG 1777
Db 215 ProGlyValHisValIlyTyrIlyLeuSerIlyProAspAspAspProLeuHisIlyGlnIlyPro 234
OY 1778 CGCTTGTGGCTAGAGATGAGAGCGAGCGAGCTGCGCGCGGATGTATGTTCTCCAGAG 1837
Db 235 ArgPheThrAlaSerMetMetGlnAlaIle-----Lys 245
OY 1838 ATCTTCATTTCCACAGCGCGCGGATGTGGGCTTACAGGCGATGATCTACAGCGCGCG 1897
Db 246 IlePheHisPheHisThrArgSerAspAlaArgLeuTyrGlnIlyLysIleTyrLysProHis 265
OY 1898 GCCTTGCAGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957

```


Db	266	AlaLeuGlnProGlyLysLysHisProIlnValLeuRheValTyrGlyProIlnVal	285
QY	1958	CAGTGGTGAATAACTCTCTTCAAAAGCATCAAGTACTGCGGCTCAACACACTGGGCTCC	2017
Db	286	GlnIleuValAsnAsnSerPheLysGlyIleTyrLysLeuArgLeuAsnThrLeuLaser	305
QY	2018	CTGGGCTACACCCGCGGTGTGATTGAACGAGGAGGCTCTGTACAGAGGCTTGCGGTTC	2077
Db	306	LeuGlyTyrAlaValValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPhe	325
QY	2078	GAAAGGAGCCCTGAAAAACAAATGGCCAGCTGGAGATGAGAGACCCAGTGGAGGCTGTG	2137
Db	326	GlnGlyAlaLeuLysAsnGlnMetGlyGlnValGlnIleGlnAspArgIlnValGlnGlyLeu	345
QY	2138	CAGTTCGTGGCGCAGAGTATGGCTTCATCAAGCTAGAGCCAGTGGCATCCATGGCTGG	2197
Db	346	GlnPheValAlaGlnLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTyr	365
QY	2198	TCCATACGGGGGCTCTCTCTGCTGCATGGGGGCTATCCAAACGCCAGGTGTTCAGGTG	2257
Db	366	SerTyrGlyLysPheLeuSerLeuMetGlyLeuIleHisLysArgGlnValRheLysVal	385
QY	2258	GCCATGCCGGGGTCCCGCGGTCACCGCTGTGATGGCTACGACACAGGTGACTAGCGC	2317
Db	386	AlaIleAlaGlyAlaProValIlnThrValTyrMetAlaTyrAspThrGlyTyrThrGlnArg	405
QY	2318	TACATGACACTCCCTGAGACACACAGACAGGGTATGAGCGGGTTCGTCGGCTCCGTGAC	2377
Db	406	TyrMetAspArgAlProGlnAsnAsnGlnIlnGlyTyrGlnIlnGlySerValAlaLeuHis	425
QY	2378	GTCGAGAACTGCCCAATGAGCCCAACCGCTGTATCTCTCAACGGCTCCCTGGAGAA	2437
Db	426	ValGlnLysLeuProAsnGlnProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGln	445
QY	2438	AAGCTGACATTTTTCACACAAATCTCTGCTCCACATGATCCAGACAGGAAAGCT	2497
Db	446	AsnValIlnSpherHeHisThrAsnPheLeuValSerGlnIleuIleArgAlaGlyLysPro	465
QY	2498	TACCAGCTC-----CAGATCTACCCCAACAGAGACAC	2530
Db	466	TyrIlnGlnValAlaIleAlaProProValSerProGlnIleTyrProAsnGlnArgHis	485
QY	2531	AGATTGGCTGCCCCGAGTGGGGCGAGACACTATGAAGTCACTGTGCTGACTTTTCAAG	2590
Db	486	SerIleArgCysProGlnSerGlyLlnHisTyrGlnValThrLeuIleuHisPheLeuGln	505
QY	2591	GAATACCTC 2599	
Db	506	GlnTyrLeu 508	
RESULT 5			
Q9BVR3	ID	PRELIMINARY:	PRT: 439 AA.
AC	Q9BVR3;		
RC	01-JUN-2001 (TrEMBLrel. 17, Created)		
RL	Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: BC000970; AAH00970.1; -		
DR	MEROPS: S09.019; -		
DR	InterPro: IPR001375; Peptidase_S9.		
DR	InterPro: IPR000379; Ser_estcs.site.		
DR	Pfam: PF00326; Peptidase_S9; 1.		

KM	Hypothetical protein.	
FT	NON_TER 1	
SQ	SEQUENCE 439 AA; 49926 MW; A18BBA9E12092BAF CRC64;	
 Alignment Scores:		
Pred. No.: 1	28e-150	Length: 439
Score: 2383.00	Matches: 439	
Percent Similarity: 100.00%	Conservative: 0	
Best Local Similarity: 100.00%	Mismatches: 0	
Query Match: 49.70%	Indels: 0	
DB: 4	Gaps: 0	
 US-09-976-674 -4 (1-2617) x Q9BVR3 (1-439)		
QY	1283 GACACTCTTCTATCCCTTCCTCCCCCAATCAGAGGGGAGAGGAGAGAGCTGTCTTTCTCCGCCGC	1342ATCCCGCC
Db	1 AspliePheTYRProPheProGlnSerGluGlyAlaSpGluIleucSPheLeuAraGla	20
QY	1343 AATGAATCAGAGCCGGCTTGCCATTGTACAAGTACACCGCGCTTTAAATACCAG	1402CCAG
Db	21 AsnGluCySylStrGlyPheCyHisIleTyLysValThrAlaValIleuYSerGln	40
QY	1403 GGCTACGATTGGAGTAGAGCCCTTAGCCCCGGGGAAGATGAATTAAAGTCCCATAAG	1462TAAG
Db	41 GlyTRAspTrpSerGIuProPheSerProGlyGlusApGIuPhelyScySProlleYS	60
QY	1463 GAAGAGATTGCTCGACACAGGGGMAATGGGAGAGTTTGGCCAGAGCAGCGGCCCAAGATC	1522ZATC
Db	61 GluGluIleAlaLeuThrSerGlyGluTrpGluValIleuAlaArgHisIleSylSerYSylle	80
QY	1523 TGGGTCAATGAGAGAACCAAGCTGTGTACTTCTCAGGGCACCAAGACAGCGCGCTGGAG	1582ZAG
Db	81 TrpValAsnGluGluThrLysLeuValTyLysPheGlnGlyThrLysAspThrProLeuGlu	100
QY	1583 CACCAACCTCTACSGTGGTACAGTATGAGACGGCGCGGGAATGCTACGCTTCAACAGCCGC	1642ZATC
Db	101 HisHisLeuTyLValValSerTYTGluAlaAlaGlyIleValArgLeuThrPro	120
QY	1643 GAGCTTCTCCATACCTGCTCCATGAGGCCAAGATTGACATGTTCTGTCAGCATACAGC	1702ZATC
Db	121 GlyPheSerHisSerCySerMetSerGlnAsnPhaSerPheValSerHisItySer	140
QY	1703 AGCGTGAACACCGCGCGCTGCGTGCACAGTTCACAGCTAGAGCGGCCGACAGACACCC	1762ZATC
Db	141 SerValSerThrProProCySValHisValTyLysLeuSerGIuProAspAspAspro	160
QY	1763 CTGCACAGCAGACCCCGCTTGCGGCTACAGATGAGAGGACAGCCAGCTGCCCGCGAT	1822ZATC
Db	161 LeuHisLysGlnProAlaGlyPheThrAlaSerMetIleAlaAlaSerCySProProAsp	180
QY	1823 TATGTTCTCTCAGAAGATTCTTCATTTCCACACGCGCTGGATGTGGCGTCTACGGCATG	1882ZATG
Db	181 TyLValProProGluIlePheHisPheHisIthrArgSerAspValArgLeuTyGI Met	200
QY	1883 ATCTTCAAGGCCSCAGCGCTTGACACCGGGAAGAAGACACCGCCAGCTCTTTTATAT	1942ZATAT
Db	201 IleTyLysProHisSalAleuGlnProGlyLysLysHisProThrValIleuPheValTyL	220
QY	1943 GGAGGCCCCAGAGTGCAGCTGTGTGAATTAATCTCTTCAAAGGCATCAATGACTTGGGCTC	2002ZATC
Db	221 GlyLysProGlnValGlnIleuValAsnAsnSerPheYselYlleYSylTyLeuAraGlu	240
QY	2003 AACACACTGGGCTCCCTTGGGCTACGGCGGTGTGATTGACGGCAGGGGCTCTGTCTAG	2062ZATG
Db	241 AsnThrIleuAlaSerLeuGlyTyLAlaValAlaValAlleasPelYarGelYserCySgln	260
QY	2063 CGAGGCGCTTCGTTCSGAAGGGCGCCCTGAANAACCAATGGCGCAGGTGAGATTCAGAGNC	2122ZATC
Db	261 ArgGIuLeuAryPheGlnGlyAlaIleuLYsAsnGImetCIuYlValGIuIleGluAsp	280
QY	2123 CAGGTGAGGGCTCTCAGATTGTGTGGCCGAGAAATATAGCTTCATCGACTGAGCCGAGTT	2182ZATG
Db	281 GlnValGIuGlyLeuGlnPheValAlaGlnIleTyTGlyGIYPheIleAspLeuSerAryVal	300

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 12:03:27 : Search time 23 Seconds
(without alignments)
3607.131 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646

Sequence: 1 MATTTGTTATRGDAATDDP.....CPESGEHYEVTLHFLQEYL 863

Scoring table:

BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	934.5	20.1	931	2	T32919
2	805.5	17.3	738	2	A87516
3	774	16.7	741	2	UC5142
4	585.5	12.6	711	2	S66261
5	508.5	10.9	793	2	T41703
6	487.5	10.5	766	1	CDHU26
7	484.5	10.4	792	1	A39914
8	480	10.3	760	1	S23752
9	477	10.3	779	2	T25174
10	467	10.1	779	2	T25173
11	445.5	9.6	803	2	I68600
12	445.5	9.6	865	2	I54331
13	442.5	9.5	803	2	A41793
14	432.5	9.3	818	1	A30107
15	417	9.0	931	2	A49737
16	394.5	8.5	795	2	F82858
17	333.5	7.2	829	2	T19514
18	313.5	6.7	759	2	I38593
19	286.5	6.2	683	2	E87495
20	270.5	5.8	709	2	B82580
21	262	5.6	743	2	I37700
22	249	5.4	657	2	E70025
23	240.5	5.2	642	2	C71137
24	234	5.0	631	2	H75007
25	215.5	4.6	732	1	UC4655
26	212.5	4.6	732	1	S07624
27	204	4.4	569	2	S74053
28	203	4.4	622	2	F71174
29	202	4.3	591	2	H72474

30	194.5	4.2	732	1	JU0132	acylaminoacyl-pept
31	194	4.2	536	2	F90299	acylaminoacyl-pept
32	191.5	4.1	582	2	D72636	probable acylamino
33	191	4.1	606	2	T35378	probable acylamino
34	191	4.1	761	2	S44807	probable acylamino
35	189	4.1	632	2	E75057	peptidase PAB1418
36	187.5	4.0	629	2	T15945	hypothetical prote
37	186.5	4.0	667	2	A87711	prolyl oligopeptid
38	185	4.0	674	2	B84381	acylaminoacyl-pept
39	180.5	3.9	608	2	F83397	probable peptidase
40	178.5	3.8	655	2	T35975	probable acyl-pept
41	176.5	3.8	819	2	B87580	conserved hypothet
42	174.5	3.8	572	2	F72455	probable acylamino
43	173.5	3.7	955	2	F84914	hypothetical prote
44	167	3.6	265	2	B84063	hypothetical prote
45	165.5	3.6	705	2	JX0194	prolyl oligopeptid

ALIGNMENTS

RESULT 1									
T32919									
hypothetical protein K02F2.1 - Caenorhabditis elegans									
C:Species: Caenorhabditis elegans									
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999									
C:Accession: T32919									
R:Maggi, L.; Goela, D.									
A:Submitted to the EMBL Data Library, January 1998									
A:Description: The sequence of C. elegans cosmid K02F2.									
A:Reference number: Z21246									
A:Accession: T32919									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: DNA									
A:Residues: 1-931 <MAG>									
A:Cross-references: EMBL:AF043699; PIDN:AA97564.1; GSPDB:GN00019; CESP:K02F2.1									
A:Experimental source: strain Bristol N2; clone K02F2									
C:Genetics:									
A:Gene: CESP:K02F2.1									
A:Map position: 1									
A:Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2;									
Query Match									
Best local similarity 20.1%; Score 934.5; DB 2; Length 931;									
Matches 264; Conservative 142; Mismatches 322; Indels 167; Gaps 33;									
QY	72	IYLLGMPYGSRENSLLYSEIPKVV-----RKPAALLLSWKOMLDHFQATPHHGVS	123						
DB	82	MAIISVGTNTQSIIFSIVTIPLELYEKAQVADRKELEKSGYNDSTIRMSCRTTPSA	141						
QY	124	EEELIRERRKLGVF-GITSYDFHSESGLELFOASNSLFH-----	161						
DB	142	EFTLQECRQSRQVYVIGISDYEL--RNGKMILMAGDQLFRYPNPLNALAIPVAPDDQS	199						
QY	162	-----CRDGGKNGFWSPMKPLEIK--TQCSGRPMDP-----K	192						
DB	200	TEPMDISGSGITSQKGSNEAPQSSVTPVPRIPKPTTSTKTPATAPPTNNFVSSAK	259						
QY	193	ICPADPAFFSFTNNSDLWANITETGEERLTPCHOGLSNVLDDPKSAGVATFVIOEEDR	252						
DB	260	VCPAASSLAIYLVNKOYV-----EKNGKIHTTSSN--SKHTINGVPSIYVEELER	310						
QY	253	FTGYWMCPTASWESGELTLRLIYEEVDESVEV---IHWSPALREKRTDSRYRPT	308						
DB	311	FEGIWW-----SES-KT-RLIYEHVNEEKVAESQFCVNDPPVA-----PKKYPRA	354						
QY	309	GSKNKIKLKLAEFGQTDQSGKTVSQEKLQVPSFLPKVYTIARAGTTRDGKXANMF	368						
DB	355	GKKNAVSTLRWVILE---NGKAYDVPLKDEV---IYKHCFEYITIRAGFSGGTTWVOV	409						
QY	369	IDRPQWQLOLVLLPPLALFIPSTE-----NEEQRLASARAVPR--	405						
DB	410	MSRDQAQCSLLIPLYTDFLLPEELGSGIKEDNLQJLSTDLNMGVMDKSHETMEKPRPKX	469						

```

0Y 406 --NNOPRVVVEEVLNWINVINDJFEYRPOSEGEDEC-FLRANECKTGSCJLYKTAVAL 461
Db 470 LRGVQ---HKRANRWINTNNAIYLKLTIDDEHPMEFTIYCLEKPNGSC-LALISAEL 525
0Y 462 KSQGYDMSPPSPGDEFECPKKEEIALTSGSEVRLARHGSKTWMNETKLVYFGQTKDT 521
Db 526 DONGY-----CRHTEQKLLMENEFSINKSMG--IYDREVLRELYVYVANNESH 569
0Y 522 PLEHLILVYVSEANGELVRLITTPFHSICMSQN-----FDM---FVSHISVSVPRCV 572
Db 570 PLEWIT-CVSHYRTGQAHQLTESI---CRSEANCKLALDDHGFACMTSVGSPARC 625
0Y 573 HVYKLSGEDDDPLHKOPRFMASSMEASCP--PD--VYPEIFHFNR-RSDVRLYGMIK 627
Db 626 RFTSRMKMENVL--PSTVVAANTTVSGHRCQORLNDHSDPEMIEFQSKTGLMHNHYMILR 683
0Y 628 PHALOPGKHPTVLVYVGPOVOLVNNSEFGIKYLRNTLSTAGYAVVYIDGSGQORG 687
Db 684 PSNDFPYKKYEVFVYVGGPDIQIIVHNDFSMIOYIR--FCRIGYVVFIDNKGSAHNGI 740
0Y 688 RFEGALKNOMQOVELERPOVELOPVAKRY-GEIDLSPVALHNGMSGGSEFLSMGLHNKPOV 746
Db 741 EFERHIIKKMSTVELEQVDESLQMLAERTGSGFMDMSRVVNHGMSYGGYMLQMIANHPNT 800
0Y 747 FKVALAGAPVTVMNAVDTGYTERIMADVPENNOMQOEGASVALYVEKJLPNEBNITLLHGF 806
Db 801 YRAALAGAVSDMRULDYPAETERMGYP-LEENHYGASSITGLYVEKLPDEPNRLMLVHGI 859
0Y 807 LDEWVNHFNHTFLVSOLIRAGKRYQOLOIIPNERHISICPESGHEVYTLHLHFOE 861
Db 860 MDENVNHHLTHLVDECKKGMKHELVIFPERRGVGNNDASITYLARMYVFAOO 914

```

```

RESULT 2
A:87516
dipeptidyl peptidase IV [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87516
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Jamb, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87516
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-738 <SMO>
A:Cross-references: GB:AE005673; NID:913423647; PIDN:AAK24125.1; GSPDB:GN00148
C:Genetics:
C:Gene: CC2154

```

	Query Match	17.3%	Score	805.5	DB 2:	Length	738:
	Best Local Similarity	28.8%	Pred.	No. 6	8e-54:		
	Matches	227:	Conservative	118:	Mismatches	300:	Indels 135; Gaps 25:
Qy	67	PHSHRLYL-CMPYSGRENSLLYSET---PKVKREALLLLSWKOMIDFHQATPRHHGY	121				
Db	47	PDGKVTVTLKGPEAANIODLMAADVGGEPRYLIDSALSSGDKEI-----	93				
Qy	122	SREELLERKRKLGVFGITSYDFHSESGFLFQAUNSLEPHCDGCKNGPMVSPMKPLEIK	181				
Db	94	SEAEARREBRANVSARGIVEYSMDRQGRTILVPLGGDL--LDAAVDG-----KITLL	144				
Qy	182	TQCSPRMDDPKICPADPAFFSFFINNSDMLVMANIETGEERRLLFCHQGLSNVLDPKSACV	241				
Db	145	TETPDEVDADAKSPKG-GVSVYRDQNLIYIKPVAGGAETALLTTDDC-----DALSFcv	196				
Qy	242	ATFVQIEERFDRTGGMWCPPTASMBESGLTKLRILEEYDESEVEVIHVPSALEERKD	301				
Db	197	AETVQEELDRFTGWMSPPDES-----RIYTRDSEGIIVLPADIGPGATVV	246				

OY	302	SYRPRGRSGKNPKIALKALAEQDSDOKIYS70---	EKELVOPFSSLFKPEVYIARAGMT	358		
Db	247	NORVPRAGRPAVVDLEFVRDL---	ASQKVALDLGANKDI-----YVARVAMS	291		
OY	359	RDGKYAMAFEDRPOQWLOJVLPLRALFIFISTENEBOJLASA	PAVRYNOQPYVVEEYJN	418		
Db	292	ADGKTVYVQRLSRQOKLDDLAPDA-----	IGAGKTIILTDPDHPFI--EVSN	337		
OY	419	WVINVDHIFYPPOSEGEDELCEFLRANECKTGFCHLYKTV	AVLKSQGYDMSFPFSDE	478		
Db	338	DE-----RPLTDC-----	TFLMGE--KDGQNHLYRYAA-----DG-----	366		
OY	479	EKCPKEEIALTSEMEVYLARHSGKIYVNEETLYVPOGKTP	RLPHNLHYVSYEAGEI	538		
Db	367	-----KLTAQITKDKDMPVIGLEG---	VDARVVAIFSAIDPIERRLYEVSYAKPGP	417		
OY	539	VLRTPGFHSCHSCSQONFDMFVSHYSVSVP	PCVNHYYKLSGP-----DDDPHLHKOPRMA	593		
Db	418	KALTSAGGWMMAAKKADNDGCAFGAGTISDPKTP	SOTALYHSA	DGKKVYRTEENKLAEGHPYR	477	
OY	594	SMMEASCPDPYVPELHFHFTRS	DVRLXGMILYKPHALOR	KKHNPVLEFVGGPOVOLVN	653	
Db	478	---YAANLP---	QREGSLKADGETLHYELKPIG	FORPAKKYRAIVS	VGGPRAHQRYM	530
OY	654	NSFKGIKYLRNLTA	LSGIYAVVYIDGSGSQRLRE	BGALKMOMQVLEDOVEGLOFPA	713	
Db	531	KNMHSPPSE---	RTYLEAGUYVIFKLDN	GSNBRSAKFMRLDKRLG	IVEEVDDLAKFLA	587
OY	714	EKYGFIDLSRYA	IGHGWSYGFELSLMGLIHR	POYFKVAILAGAPVTVM	AYDVGTYTERYMDV	773
Db	588	SO--PYVDADK	LGVMGWSYGGFPMALMLLT	TAENTPFKGAAGAPRT	ESLQDTRATERYMK	646
OY	774	PENNOHGVEAGS	VALHWEKLPNEBRNLLI	HGFLDENVEHFTHTN	FLVSQILNAGKPYOLO	833
Db	647	PDEKNAGAYASD	INNRIKDL--APGSLL	LLHGADNDVIFENSTR	RLMALQKAILFEWA	704
OY	834	IYPNRRHS	841			
Db	705	IYPERRRHS	712			

RESULT 3
JC5142
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C:Species: Xanthomonas maltophilia
C:Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C:Accession: JC5142
R:Kobashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A:Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and express
A:Reference number: JC5142; MUID:97164011; PMID:9010738
A:Accession: JC5142
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-741 <KAB>
A:Cross-references: DDBJ:D83263; NID:g1753196; PID:BA011872.1; PID:g1753197
C:Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; membrane bound
F:4-18/Domain: transmembrane #status predicted <TM>
F:610/Active site: Ser #status predicted
F:685/Active site: Asp #status predicted
F:711/Active site: His #status predicted

Query Match	16.7%	Score	774	DB 2	Length	741			
Best Local Similarity	29.3%	Pred. No.	1,9e-51						
Matches	226	Conservative	125	Mismatches	275	Indels	146	Gaps	29
QY	108	MIDHQAAPHHCYVREDELLERKRRIGVF	GITSYDFHSEGLFLQASNS	-LFHCD	164				
DB	79	LVDSKVVLPGETTSDDEKARRRRIAMGTIVYQWSDQARLLFPFGGELYLYVDLQK	138						

```

QY 165 GKGNGVSPMKPLEIKTQCSGPRMDPKICPADAFSFINNSDLVANIETGEERRLT 224
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 EKG-----AAVRQL---THGEGFATDAKLSPEKG-GFVSFIRGRNLMTVIDLASCGRQMLT- 188
QY 225 CHQGLSNVLDPKSGAVTFVYIOEEDFRFTGWMCPITASMESEGLKRLIRIYEVDSE 284
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 -ADGSTT-----GNGIAFEVADEMDRHTGTWMAPDDS-----ALAYARIDESP 232
QY 285 VEVIHVPSPALEER-----KTD--SYRYPRTGSKNPKIATLAEFOTDSOGKIYSTQEK 336
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 VPV-----OKREYVADRDTVIEQRYPAAGDANVOYKIGVSPAEQAOTQWIDLCKE 284
QY 337 ELVOPFSSLPFKVEYIARAGWTRDQKYAMAMELDRPQOQLVLLPPLFTIPSTENEOR 396
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 QDI-----YLAVNM--RDPQHLSFQKSRDQKKLDV-----EVT 318
QY 397 LASARAVRNQPYVYVEEVNINWITVPPPOSEGEDELCLFRANECKTGCHLYK 456
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 LASN-----QORVLAHETSFPTWPLHN-----SLRFLLDGSILMSE--RTGFQHDYR 364
QY 457 VTAVLKSQGYDMSEPSGDEDFKCIKEETALTSGEM---EVLARHSGKIWNVEETKLV 513
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 I-----DSRG-----KAAALTHGNMSVDLLA-----VDEKAGLA 394
QY 514 YFOGTRDPLEHHLVYVEAAGEIVRLTPGFSHSCMSQNFDMFVSHYSSVSTPCVH 573
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 YFRAGIESARSOIYAVLQ--CGQPORLSKAPGMHSAFARASVYVWSMNSNSTPPOTE 453
QY 574 YKLSGPD-----DDPLHKOPREWASMEASCPDDVYPRFIHFHTRSDVR--LY 622
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 LFRANGEEIATLVENDLADPKHRYARY---REAOR-----PVEFGTLTAADKTEPLN 502
QY 623 GMIVPHALOPGKKHPTVLFYGGPOVOLYNNSEKGI--KYLRLNLTSLGTVAVVYIDGG 681
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 503 YSVIKPAFDPDKRIPEVAVYVIGGPASQTVTDSWPGRGDHLNÖYLLAQGGYVFSLDNG 562
QY 682 SCQGRLEEGALKNQMGVEIEDOYEGLOFYAEKGYFTDLSRVAITHGMSGYGLSLMGLI 741
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 563 TPRGRDRGAGALYKGGQVEVADQRGVAMLKQ--PWVDPARIGVQSGNGGYMLMLLA 621
QY 742 HKPOYFKVAGAPYVVMAYDTGTERMDVPENNQHGEGSVALHVEKLPNEPNRL 801
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 KASDSYACVAGAPATDGLYDSHYTERYMDLPARNDAGYREARVLTHEGLRSP---LL 678
QY 802 ILHGLDENVHEFTNFVLSOLIRAGKPYOQIYPERNSHRCPSGEHYEV 853
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 679 LIHGADNVLFTNSTLSMSALQKRGQPELMTYPAKKGHLSGADALHRYRV 730

RESULT 4
S66261
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
C:Species: Flavobacterium meningosepticum
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S66261
R:Kasahima, T.; Yoshida, T.; Ito, K.; Yoshimoto, T.
Arch. Biochem. Biophys. 320, 123-128, 1995
A:Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from F.
A:Reference number: S66261; MUID:95314307; PMID:7793970
A:Accession: S66261
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-711 <KAB>
A:Cross-references: EMBL:D42121, NID:9577283; PIDN:BA07702.1; PID:9577284
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase

Query Match 12.6%; Score 585.5; DB 2; Length 711;
Best Local Similarity 25.2%; Pred. No. 6,9e-37;
Matches 184; Conservative 130; Mismatches 266; Indels 147; Gaps 24;
QY 142 YDF-HSEGLFLQASNSLFLHCRDGGKNGFVWSPMKPLEIKTQCSGPR---MDPKICPA 196
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 81 YTFNSDESKILLQKSSQSIYRHSFLGK-----FEVKDLKSRTVSLNNANWIOEPRFSP- 134
QY 197 DPAPFSFINNSDLVANIETGERLRTFCHQGLSNVLDPKSGAVATFVIOEEDFRFTG 256
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 DSKVAFIADNNLFTYQDLNTGKITQIT--TDGKKNELIN---GLGDWYEEERGHADY 188
QY 257 WMCPTASMESEGLKRLIRIYEVEDESEVEVHNH-----SPALEERKTSYRYPRTGS 310
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 QMKNAGD-----ALVFRPDERRYPEINITYQNLYP-----KIMTYKPKAGE 233
QY 311 KNPKITALKLAEFQYDSOGKIYSTQEKELVQPSLSLPKVEYIARAGWTRDQKYAMAMP 370
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 ENSAVTAVL--YQLSS-----GRSAOLNFGS 257
QY 371 RPOQLQVLVLLPPLFIPTSTENEORLASARAVRNQPYVY-----YEEVTWV 420
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 SEKYIT-----POLF--QTNANDEIVANANRHNQKVDLLKNTKTAAVSKLFTETDNAM 310
QY 421 INVHDIFFYPPQSEGEDELCLFRANECKTGFCCHLYKVTAVLKSQGYDMSEPSGDEBK 480
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 IETDNLTFEF-----LDDNSFLMASE-RDGHRLIYWDYAGK----- 346
QY 481 CPKEEIALTSGEWEVLARHSGKIWNVEETKLVYFGTKRPLEHHLVYVEAAGEIVR 540
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 --LKKQV--SKGDWEITINYG---YNPKTKEVYIQTTEKGSINKVSKLININ-TGKTOL 397
QY 541 LTPPFHSCMSQNFDMFVSHYSSVSTPCVHYVTKLSGD-----DDPLHKOPRRA 593
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 LSNAGNNSAARSKTFENFINTSTPAKYPTKYLKDANCKDKELQNDNDLLNKK- 453
QY 594 SMWEASCPDDVYPRFIHFHTRSDVRLYGMITYKHALQPGKHPVLFYGGPOVOLYN 653
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 -----SDNITAEFTITPMAAGDQMANNAIKRKNRPDAKRYVPMFOYSGPSQOYA 505
QY 654 NSFQIKYLRILNTLASLGYAVVVDGRSGCQGRLEEGALKNQMGQVEIEDOYEGLOFYA 713
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 506 NSMDGCGNWGFMLLQKQGLVYCVDRGTGFGFGTKYKKYTKNLKYEIEDQITAAKWL 565
QY 714 EKYGFIDLSRVAITHGMSGYGLSLMGLIHKPOYFVAVIAGAVTYVMAYDQCYTERYMDV 773
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 NO-STYVDSRIGIFGWSYGGYMASLMTKGADEVFKGILAVATYVMRFDSITYTERFLQT 624
QY 774 PENNOHGEAGSVALHVEKLPNEPNRLLIHGFLENVHEFTNFVLSOLIRAGKPYOQ 833
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 625 PQENNDGDLNLSPTTYAKL---KGFLLIHGTADNHNHFOQSMSESEALLQNKQFDEM 681
QY 834 IYPERHST 842
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 682 AYPDRKHSI 690

RESULT 5
T41703
dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T41703
R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z22011
A:Accession: T41703
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-793 <MUR>
A:Cross-references: EMBL:AL031180; PIDN:CAA20138.1; GSPDB:GN00066
A:Experimental source: strain 972h-; cosmid c2E11 -chimeric
C:Genetics:
A:Gene: SPAC2E11.08
A:Map position: 1
C:Superfamily: dipeptidyl-peptidase IV

Query Match 10.9%; Score 508.5; DB 2; Length 793;
Best Local Similarity 25.4%; Pred. No. 7.5e-31;

```



```

QY 272 TLRLIVEDESEVEVH---VSPALERKTDYRYPRGSKNPRIALKLAEPOTDSOG 328
D 223 ---LAIQAFNDTEVPLEIETSYSPDSIDQYPTVTRVPFKACAVNPTV--KFFVNTDLS 277
QY 329 KIVSTOEKELVQPSSEFPKVEYIARAGWTRDGKYAMAFDRPOWLQTLPPALFIP 388
D 278 SVTATSIQITAPASMLIGD-HYICDVTWATQERIS-----LQWL----- 316
QY 389 STENEORLASARAVPRNVQYV---YEVTNVMINVDIYPPQSGEDELCLR 443
D 317 -----RRIQNYVMIDICDDYDESSGRW----- 337
QY 444 ANECKTGCHLYKTAVALKSGQYDMSEPFSPDE-----FK 480
D 338 --NCLVARQHLEMTT-----GWGRFRSEPHFTLDGSEFKIISNEGYRHCYFQ 388
QY 481 CPKEEIALTSGEWEVLARHGSKIWNBEKLYFOGTRDPLHLHYVSEYAG--- 536
D 389 IDKDCFFITFGTWEVIGIEAL-----TSD-----YLYISNEYKGMG 428
QY 537 -EIVRLTPGSHSCSQNDFMVSYSVTPPCVHYVL--SGP----- 580
D 429 RNLKYIQLIDYTKVTCSELNPERCOYYSFSEKAAKYOLRGSGPLPYTLHSSVND 488
QY 581 -----DDPLHKQPRFMAEMEAASCPDYVPEIHFHTRSDVRL-YGMIKPHALQ 632
D 489 KGLRVEDNSALDK-----MLQWQ-----MPSKIDFTILLETWKWYMLPPH-FD 535
QY 633 PGKHPYVLFYVGGPOVOLVNNSEFKGIKYLRLN--TLASIGYAVV--IDGRSGCORGLR 688
D 536 KSKYPLLDVYAPGSCQKADTVF-----RLNMATYLASTENIIVASFDRSGSYQDK 589
QY 689 FEGLKMQMGVEIEDOYEGE-QVYAEYGGTIDSRVAIHGWSIGGELISGLHKPOVF 747
D 590 IMHAINRRLGFEVEDOIEAARQF--SKMGFVDMKRIAIMGWSYGGYTSVVLGSGSGVF 647
QY 748 KVALAGAPVYVMAVDYGTERTYMDV--PENNGYEGASVALHVEKLPNPNRLILIHG 805
D 648 KCGIAPVAVSWMEYDYSTRYKMGLPPEPNLDHTRNISTVMSRAEN--KOVETLLIHG 705
QY 806 FLDENVHEFTNPLVSOLIRAGKPYOLOIYENERSIRCPSENGEHEVETLLFLQ 861
D 706 TADNVHFGQSAQISKALVDYGVDFQAMWMTYDEDHGIASSTAHQIITYHMSHFIRQ 761

RESULT 7
A39914
dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
N:Alternate names: GP110; membrane glycoprotein 110K; OX-61
N:Contains: dipeptidyl-peptidase IV, soluble form
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
C:Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781
R:Hong, W.; Doyle, D.
Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987
A:Title: cDNA cloning for a bile canalicular domain-specific membrane glycoprotein of rat
A:Reference number: A39914; MUID:88068516; PMID:3479775
A:Accession: A39914
A:Molecule type: mRNA
A:Residues: 1-792 <HON>
A:Cross-references: GB:J02997; NID:g204463; PIDN:AAA41272.1; PID:g204464
R:Ogata, S.; Miumi, Y.; Ikehara, Y.
J. Biol. Chem. 264, 3596-3601, 1989
A:Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA at
A:Reference number: A33315; MUID:89123496; PMID:2563382
A:Accession: A33315
A:Molecule type: mRNA
A:Residues: 1-37, 'A', 39-182, 'I', 184-331, 'T', 333-351, 'C', 353-393, 'V', 395-561, 'L', 563-623,
A:Cross-references: GB:J04591; NID:g203973; PIDN:AAA41096.1; PID:g203974
A:Note: the authors translated the codon GCG for residue 38 as Arg, ACC for residue 332
A:Accession: B33315
A:Molecule type: protein

```

```

A:Residues: 1-20:35-54;427-443:505-509:511-520:530-538:593-600:602-608:618-627 <H02>
R:McCaughan, G.W.; Mickson, J.E.; Crewick, P.F.; Correll, M.D.
Hepatology 11, 534-544, 1990
A:Title: Identification of the bile canalicular cell surface molecule GP110 as the ec
quence.
A:Reference number: A60730; MUID:90228896; PMID:1970322
A:Accession: A60730
A:Molecule type: protein
A:Residues: 28-47, 'XX', 50-53, 55-58 <MCC>
R:Ogata, S.; Miumi, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.
Biochemistry 31, 2582-2587, 1992
A:Title: Identification of the active site residues in dipeptidyl peptidase IV by aff
A:Reference number: A42203; MUID:92190188; PMID:1347701
A:Accession: A42203
A:Molecule type: protein
A:Residues: 'R', 625-630, 'X', 632-648 <OG2>
R:Iwaki-Egawa, S.; Matanabe, Y.; Fujimoto, Y.
Biol. Chem. Hoppe-Seyler 374, 973-975, 1993
A:Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptid
A:Reference number: S38949; MUID:94128239; PMID:7905271
A:Accession: S38949
A:Status: preliminary
A:Molecule type: protein
A:Residues: 281-302 <IWA>
R:Hong, W.; Doyle, D.
J. Biol. Chem. 263, 16892-16898, 1988
A:Title: Membrane orientation of rat gp110 as studied by in vitro translation.
A:Reference number: A31781; MUID:89034185; PMID:3162821
A:Accession: A31781
A:Molecule type: mRNA
A:Residues: 1-40 <H03>
A:Comment: This protein is localized to the bile canalicular, which is the apical dom
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidyl-peptidase hydrolase, glycoprotein, homodimer, liver, serine prote
F:1-792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATM
F:1-6/Domain: intracellular #status predicted <INT>
F:7-28/Domain: transmembrane #status predicted <TMN>
F:29-792/Domain: extracellular #status predicted <EXT>
F:29-34/Domain: propeptide #link MATS #status experimental <PRO>
F:35-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MATM>
F:83,90,148,217,227,319,521,686/Binding site: carbohydrate (asn) (covalent) #status p
F:631/Active site: Ser #status experimental
F:709,741/Active site: Asp, His #status predicted

Query Match 10.4%; Score 484.5; DB 1; Length 792;
Best Local Similarity 23.6%; Pred. No. 5.4e-29;
Matches 205; Conservative 123; Mismatches 334; Indels 207; Gaps 36;

QY 63 DESGHSRLV---YLGMPYSGRENSILY--SEIRKYRKELALLLSKKOMLDHQAQTPH 117
D 31 DEAAADSRRTYTLADLYLNKTRFKVSYSLRWSDSEYLYKQENNTLL-----FNA--E 80
QY 118 HGVSREBELLRERKRLVFCITSYDFH-SESGLELFOASNLFLCRDGGKNGFVSPMK 176
D 81 HG-----NSIFLENSTFEIFGDISDYSVSPDRFLVLEYNVKKMRKSYASYSITDNL 136
QY 177 PLEITQCGSPRMDPKICPADPAF--FSFINNSDLWVANIETGEERRLTFCHOGLSNVLD 235
D 137 KROLITEKRIPTNNTQWIMWSEGHKLAVYVWKNNDIYKIEPHLPSHRT--STGKENVIFN 194
QY 236 PKSGVAFVIOEE-FDGRFTGYWGCPTASWSSSEDLKRLITIEVDSSEVEVH---VP 291
D 195 -----GINDVYEEETFGAISALMWSNCTF-----LAIQAFNDTEVPLEIETSY 240
QY 292 SPALERKTDYRYPRGSKNPRIALKLAEPOTDSOGKIVSTOEKELVQPSSEFPKVEY 351
D 241 DESIQYPTVWLPYKACAVNPTV--KFFIVNTDLSSTTTTIPQITAP-ASVTTGGHY 297
QY 352 IARAGWTRDGKYAMAFDRPOWLQTLVLPALFIPSTENEORLASARAVPRNVQYV 411
D 298 LCDVAVWSEDRIS-----LQWL-----RRIQNYMS 321

```

QY 412 V-----YEVNTWVNIWHDIFPPPOSEGEDELCLFLRANECKTGFCNLYKVTAVLKSGY 466
 Db 322 VMACIDDKTNLW-----NCPPTQEH-----LETSAT 349
 QY 467 DWSEPSGDEP-----KCPKEELALTSGEWEVLARH 500
 Db 350 GWMGRFRPAEPHFTSDGSSFTKYIVSKDKYKHICQFQKDRKRPEDCFTFTKGAEVLS-- 407
 QY 501 GSKIVWEETKLVYFQGT-KDTPLEHLLVYVSYEAGEIVALTTPGFSHSCSQMDF 559
 Db 408 ---LEALTSYLVYISNEYEMPGRLYKIQ-----LTDHNNKC-LSCDLNPE 453
 QY 560 VSHYSSVTPPCVHYVLS--GP-----DDPLHKQPRFASME 597
 Db 454 RCQYVSVLSKEAKYVGLGRGPGPLPYTLHRSTDKELRYLEDNSALDK-----MLQ 506
 QY 598 AAGSPRPVYPELPHFTSRSDVRL-YGMITYKPHALQPGKHPVLYFYGGGPOVOLVNSF 656
 Db 507 DVQ-----MPSKIDFLVLETRFWYQMLRPH-FDKSKYRPLLDVYAGCSQKADAP 560
 QY 657 KGIKYRLNLTLASLGAVVY-IDGRGSCQGLRFGALKNQMGVEIEDQVEGL-QEVAE 714
 Db 561 ---FENNATYLASTENITIVASFQGRSGYQGDKIMHAINRRLGLTEVEDQIEARQGL-- 615
 QY 715 KYGFIDLSRAIHGWSYGGFLSLMGLIHKPOVFKVAIAGAPVYMAVDTGTERYMDV- 773
 Db 616 KMGFVDSKQVAILMGWSYGGVYVTSVWGLSGSGVFECGIAPVPSRWEYDYETTERYMGLP 675
 QY 774 -PENNOGYEAGSVALLHVEKLPNEPNRLLIHGFLDENVHFHTNPLVSOLIRAKRYQL 832
 Db 676 TPEDNLDHYNSTVMSRAENF--KQVEYLLIHGTADNDVHFOQSAISKALVDAGVDFOA 733
 QY 833 QIYPERHNSIRCESGEHYETLLHFOE 861
 Db 734 MWYTDDEHGIASSTAHQHITSMSHFLQ 762

RESULT 8
 S23752
 dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse
 N:Alternate names: CD26 alpha subunit; THAM alpha subunit
 C:Species: Mus musculus (house mouse)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S23752; A46465; A56030
 R:Marguet, D.; Bernard, A.M.; Vivier, I.; Darmoul, D.; Naquet, P.; Pierres, M.
 J. Biol. Chem. 267, 2200-2208, 1992
 A>Title: cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-di
 A:Reference number: S23752; MIMD:92129288; PMID:1370813
 A:Accession: S23752
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-760 <MAY>
 A:Cross-references: EMBL:X58384
 R:Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M.;
 J. Immunol. 147, 447-454, 1991
 A>Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase
 A:Reference number: A46465; MIMD:91302787; PMID:1712807
 A:Accession: A46465
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-20 <VTY>
 A:Experimental source: M14 T thymoma cells, Swiss nu/nu
 A>Note: Sequence extracted from NCBI backbone (NCBI:42236)
 R:Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.
 Biochemistry 33, 15204-15214, 1994
 A>Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.
 A:Reference number: A56030; MIMD:95092780; PMID:7999781
 A:Accession: A56030
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 746-760 <BER>
 A:Cross-references: GB:U12620
 C:Genetics:

A:Gene: CD26
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
 F:213,223,315,514,679/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:624,702,734/Active site: Ser, Asp, His #status predicted

Query Match 10.3%; Score 480; DB 1; Length 760;
 Best Local Similarity 24.6%; Pred. No. 1,1e-28;
 Matches 214; Conservative 113; Mismatches 326; Indels 218; Gaps 39;

QY 63 DESPHSHRLX-----YGMPTGSRNSLTY-SELPKVKREALLLSWKQMLHFOATPH 117
 Db 31 DEAAADSRRTYSIADYIKSTFRVKSYSLSLMWVSDPEYLYKQENNTLL-----LNAE 80
 QY 118 HGYSRREELLREKRRLGVGITSDFHSES--GLTFQASNSLPHCRDGGKNGFWSPM 175
 Db 81 HG-----NSSTFLENSTFESG-----YHVSPLDLFVLENTYKQKRHSYASTYNTYDV 131
 QY 176 KPLEIKTQSGPRMDPKICPADPAE-FSFINSDLMVANTIE-----TGEERLTF 224
 Db 132 NKROLITEERKIPNNTQWITWSPGHNKLAYWKNDIYV-KVEPHLSHRITSTGEE----- 185
 QY 225 CHQGLSNVLDPRKSAGVATVIOE-PDRFTGYWMCPTASMESEGLTLRLIYEVEDS 283
 Db 186 -----NVIYN---GITDWYEEVEVGAVALMWSPPNNTF-----LAYAQFNDT 225
 QY 284 EVEYIH---VPSPALERKTDYSRYPRTGSKNPKIALKLAFOGDSQGIYSTQEKELVQ 340
 Db 226 GVPRLIEYSFSDLSQYKRWIPIPKAGAVNPYKFFIYNIDLS-- 272
 QY 341 PFSSLPKVEYIARAGWTRDQKY---AMAFIDRPOOMLOLVLLPALFPSTENEOR 396
 Db 273 -SSSAPIQIPAPASVARGDHYLCADVMAETEIRISQWL----- 310
 QY 397 LASRAVRNVRQYVY-----YEEV-----TNWVNIWHDIFPPPOSEG 435
 Db 311 -----RRQNTSVNACIDYDKINLTWNCPSQOHNEMSTGTGWRFRPAEPHFTSDG 362
 QY 436 EDELCLFLRANECKTGF---CHLYKVTAVLKSQGYDSEPSGDEPKCPKIEEIALTSG 492
 Db 363 SS---FYKIISDKQGYKHICHFR-----DKKDCIF-----ITKG 394
 QY 493 EMEYLAHNGSKIVNEETKLVYFQGT-KDTPLEHLLVYVSYEAGEIVRLTTPGFSHSCS 551
 Db 395 AWEYIS-----IEALTSYLVYISNQYKEMPGCNLYKIQLTDTHTNVKCL-----SCD 442
 QY 552 MSQ-----NEDMFVSHSVSFTPCVHYVYKLSGDDPLRHKORPFMASMMEASG-- 601
 Db 443 LNEPRQYVAVSFSSKEAKYQLOGCWGPGPLPYTLHRST--HKELR-----VLEDNSALD 495
 QY 602 ---PPDYVPELPHFTSRSDVRL-YGMITYKPHALQPGKHPVLYFYGGGPOVOLVNSFK 657
 Db 496 RMLQDVQMPKIDFLVLETRFWYQMLRPH-FDKSKYRPLLDVYAGCSQKADASF 553
 QY 658 GIKYRLN---TLASLGAVVY-IDGRGSCQGLRFGALKNQMGVEIEDQVEGL-QEVAE 712
 Db 554 -----RLNMAITYLASTENITIVASFQGRSGYQGDKIMHAINRRLGLTEVEDQIEARQV 608
 QY 713 AEKYGFIDLSRAIHGWSYGGFLSLMGLIHKPOVFKVAIAGAPVYMAVDTGTERYMDV 771
 Db 609 ---KMGFVDSKQVAILMGWSYGGVYVTSVWGLSGSGVFECGIAPVPSRWEYDYETTERYMG 666
 QY 772 -DVPENNQGYEAGSVALLHVEKLPNEPNRLLIHGFLDENVHFHTNPLVSOLIRAKRY 830
 Db 667 LPIPEDNLDHYNSTVMSRAENF--KQVEYLLIHGTADNDVHFOQSAISKALVDAGVDVF 724
 QY 831 QLOITPERHNSIRCESGEHYETLLHFOE 861
 Db 725 QAMWYTDDEHGIASSTAHQHITSMSHFLQ 755

RESULT 9
 T25174

[illegible]

QY 534 AAGEIVRLTTPGFSHSCSM-----SONFDMFYSHSYSTPCVH--VYKLSGPPDDP 584

Db 444 LADL-----SRNSAYCISCSIKNCTWAQAOAMDQMTAIVSCCKGPAAPHTAIYNLTRMDSK 500
QY 585 LHKQ-----PRFASMAEASCPDDYVPELFFHHTSRDVLGYMKYKPHALQCKKH-- 637
Db 501 KTEHANLLYDTTYONRVEEGLP--VLIKETIKISDPDALIKLST--PKDIYVRDHQA 556
QY 638 -PVLFFYGGPOVOLLVNNSEK-GIKYLRMLNTLASGLYAVV--IDGRSGCGRLEFEGALK 694
Db 557 IPLVHYGGNDQNTKEATQIGIE----EYVASASQAAILRIDGRSGGGMKRYSAHY 612
QY 695 NQMGVEIEDVQEGLOFAEYK-GEIDLSRYAHGWSYGGFLSLMGLHKP-QVFKYAIA 752
Db 613 GQLCTVEVEDIKAIKYVLRRLYRHLIDARVAVFGWSYGGFMTLSMVEAEQEFKCAVS 672
QY 753 GAPYVMAVDQGYTERYM-DVPENNQGYEAGSVALHVEKLPN-EPNRLLILGHFDEN 810
Db 673 VAPYTNFAYIDATYTERYMGDAP-----LESTYSDYTKLDFNKSTRLLMHGLDDN 724
QY 811 VHFHTNFVLSQILIRAGKPYOLQIYPERHSIRCPESGEHYEVTLLHFLQE 861
Db 725 VHFQNSAILDELQNRGYDFLWYPPNQAHSLSRTS--HYVGKMTHTFLRQ 773

RESULT 11
168600
dipeptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C:Accession: I68600
R:Yokocani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat
A:Reference number: I54331; MUID:93372805; PMID:8103397
A:Accession: I68600
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-803 <RES>
A:Cross-references: GB:M96860; NID:g306707; PIDN:AAA35761.1; PID:g306708
C:Superfamily: dipeptidyl-peptidase IV

Query Match 9.68; Score 445.5; DB 2; Length 803;
Best Local Similarity 23.88; Pred. No. 5,7e-26;
Matches 176; Conservative 120; Mismatches 268; Indels 175; Gaps 31;
QY 150 LFLQASNSLFCHRDGNGFMVSPMKPLEIKTCGSRMDPKICPADPAFFSFINSDL 209
Db 193 IFIRE--NNIYYCAHVQKQAIRV----- 213
QY 210 WVAНИЕGEERRLTFCHQGLSNVLDPKSAGVATFVIOEEDRFTGYWMCPTASMEGSEG 269
Db 214 ----VSTGKEGVI---YNGLSDWLYE-----EILKTHIAHWSPDGT----- 249
QY 270 LKTRILYEEDSEVEYIHWPS-----PALERKTDSTYPRGTGSKNPYIAKLAIEFQ 323
Db 250 ----RLAYAAINDSRVPMELPTGTSTYPTVK-----PYHYPRAGSNPISLHV1--- 297
QY 324 TDSQGIKISTQEKILVOPFSSLPFKVE--YIARAGWTRDGYAAMAFDRPQOMLOLYL 381
Db 298 ----GLNGPTHLDEMPPDD--PRMREYITTYMKATSTKVA--VTWLNKQNNVSIILFLC 349
QY 382 PPALEIFSTENEGRLASARAVPRNVPYVVEEVTNWI---NVHDIFFYFPQSEGEDE 438
Db 350 DATGVCTKKHED-----SEAWLHRQNDPEYF-----SKDGRK 383
QY 439 LCFIRANECKTGFCILYKVTAVLKSQGYDWSPEPSGDEKCPKIEIALTSGMEVLA 498
Db 384 FFFIRAIIP--QCGRKCKFYHIV-----SSSQPNSSNDN-----IQ---SITSGDMV-- 425
QY 499 RHGSKIYW--NEETKLYVFOGKTPTLEHHLVYVSEYEAAGELIVRLTTPGFSH---SCMSQ 554
Db 426 ---TKLIAYDEKGNKIYFLSTEDLPRRROLYSANTEG-----NFNQCLSCDLVE 472

QY 555 NFDMEVSHYSVSPPCVHYVYKLSGPD-----DDPLHKQPRFWASME--AASCPDYVP 607
Db 473 NCTEYASFSHSMQ---FLLKCEGPGVPYVTYNTDDKKMKMFLJETNEHVKKAIANDROM 529
QY 608 PEIHFHTR--SDVRLYGMIRYKPHALQPGKKHPTVLPFYGGQVOLLVNNSEFGIKYLRANT 666
Db 530 PKVEYRIEIDYMLPMQILKPATFTDTTHYPLLVVDGTPGSSQVKEKEF-VSMETV-M 587
QY 667 LASGYAVVYIDNGSGCORGLREFGALKNMGQVEIDQVGLQFAVEKYFIDLSRAVI 726
Db 588 VSHGAVVYKCDGKSGFQGTCKLLHEVRRGLLEEKQDMFAVITML--KEYIDRTRYAV 646
QY 727 HGWSYGGFLSLMGLHK---POVEKYAIAQAPVTYMAAYDTGYTERYMDVPENNQGYE 782
Db 647 FGKDYGYLSTYILPAKENGOGFTFCGSLPITDFELKYSASERSYRLGLHGDNDAYE 706
QY 783 AGSVALVEKLPNPNRLLIHGFLDENVHFHTNFVLSQILIRAGKPYOLQIYPERHSI 842
Db 707 MTKVAHRYASAL--EEQGFLLIHPADEKIHFOHTVAELTOLIRKANYSLOIYDESHYF 764
QY 843 RCPESGEHYEVTLLHFLQE 861
Db 765 TSSSLKQHLRYSIINFYE 783

RESULT 12
154331
dipeptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: I54331
R:Yokocani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-re
A:Reference number: I54331; MUID:93372805; PMID:8103397
A:Accession: I54331
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-865 <RES>
A:Cross-references: GB:M96859; NID:g306705; PIDN:AAA35760.1; PID:g306706
C:Superfamily: dipeptidyl-peptidase IV

Query Match 9.68; Score 445.5; DB 2; Length 865;
Best Local Similarity 23.88; Pred. No. 6.4e-26;
Matches 176; Conservative 120; Mismatches 268; Indels 175; Gaps 31;
QY 150 LFLQASNSLFCHRDGNGFMVSPMKPLEIKTCGSRMDPKICPADPAFFSFINSDL 209
Db 255 IFIRE--NNIYYCAHVQKQAIRV----- 275
QY 210 WVAНИЕGEERRLTFCHQGLSNVLDPKSAGVATFVIOEEDRFTGYWMCPTASMEGSEG 269
Db 276 ----VSTGKEGVI---YNGLSDWLYE-----EILKTHIAHWSPDGT----- 311
QY 270 LKTRILYEEDSEVEYIHWPS-----PALERKTDSTYPRGTGSKNPYIAKLAIEFQ 323
Db 312 ----RLAYAAINDSRVPMELPTGTSTYPTVK-----PYHYPRAGSNPISLHV1--- 359
QY 324 TDSQGIKISTQEKILVOPFSSLPFKVE--YIARAGWTRDGYAAMAFDRPQOMLOLYL 381
Db 360 ----GLNGPTHLDEMPPDD--PRMREYITTYMKATSTKVA--VTWLNKQNNVSIILFLC 411
QY 382 PPALEIFSTENEGRLASARAVPRNVPYVVEEVTNWI---NVHDIFFYFPQSEGEDE 438
Db 412 DATGVCTKKHED-----SEAWLHRQNDPEYF-----SKDGRK 445
QY 439 LCFIRANECKTGFCILYKVTAVLKSQGYDWSPEPSGDEKCPKIEIALTSGMEVLA 498
Db 446 FFFIRAIIP--QCGRKCKFYHIV-----SSSQPNSSNDN-----IQ---SITSGDMV-- 487
QY 499 RHGSKIYW--NEETKLYVFOGKTPTLEHHLVYVSEYEAAGELIVRLTTPGFSH---SCMSQ 554
Db 488 ---TKLIAYDEKGNKIYFLSTEDLPRRROLYSANTEG-----NFNQCLSCDLVE 534

Qy 811 VHFHTNPLVSQILRAG-KPYOLOIYPNERHSIR 843
|| : | | | | | : | : | | |
Db 866 VHIQNTFRLVDQNLGLTNYDMHIFPDSHSIR 899

Search completed: December 12, 2002, 12:06:37
Job time : 29 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 11:54:42 : Search time 14 Seconds
(without alignments)
2556.719 Million cell updates/sec

Title: US-09-976-674-3
Perfect score: 4646
Sequence: 1 MATTGPTADRGDAATDPD.....CPESGEHYETLLHFLQGYL 863

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	505.5	10.9	765	DPP4_BOVIN	P14425 bos taurus
2	498.5	10.7	767	DPP4_RAT	P14740 rattus norv
3	489.5	10.5	766	DPP4_HUMAN	P27487 homo sapien
4	486	10.5	765	DPP4_FELCA	O9N217 felis silve
5	481	10.4	760	DPP4_MOUSE	P28843 mus musculu
6	447.5	9.6	859	DPP6_RAT	P46101 rattus norv
7	445.5	9.6	865	DPP6_HUMAN	P42658 homo sapien
8	442.5	9.5	863	DPP6_BOVIN	P42658 bos taurus
9	438.5	9.4	761	SEPR_MOUSE	P97321 mus musculu
10	432.5	9.3	818	DAP2_YEAST	P18962 saccharomyc
11	426	9.2	760	SEPR_HUMAN	O12884 homo sapien
12	417	9.0	931	DAP1_YEAST	P33894 saccharomyc
13	249	5.4	657	YUHL_BACSD	P39839 bacillus su
14	215.5	4.6	732	ACPH_HUMAN	P13798 homo sapien
15	212.5	4.6	732	ACPH_RAT	P13676 rattus norv
16	194.5	4.2	732	ACPH_PIG	P19205 sus scrofa
17	191	4.1	761	YL31_CAREL	P34422 ctenorhadi
18	165.5	3.6	705	PPCE_FLAME	P27028 flavobacter
19	160	3.4	710	PPCE_BOVIN	O9X242 bos taurus
20	154	3.3	710	PPCE_PIG	P23687 sus scrofa
21	151.5	3.3	705	PPCE_FLAME	P27195 flavobacter
22	144	3.1	710	PPCE_MOUSE	O9QUR6 mus musculu
23	138.5	3.0	689	PPCE_AERHY	O06903 aeromonas h
24	136	2.9	710	PPCE_HUMAN	P48147 homo sapien
25	121.5	2.6	1427	ABCL1_SCHPO	O92337 schizosacch
26	118.5	2.6	726	Y4NA_RHISN	P55577 rhizobium s
27	115	2.5	436	PAFA_CAVPO	O90683 c platelate
28	115	2.5	690	PTRB_MORLA	O95936 moraxella l
29	113.5	2.4	717	MCCA_MOUSE	O9QMR6 mus musculu
30	113	2.4	857	LOX3_SOYBN	P09186 glycine max
31	112.5	2.4	861	LOX3_PEA	P09918 pisum sativ
32	112	2.4	557	SASB_ANAPL	O04791 anas platyr
33	110.5	2.4	550	CRYS_DICDI	P21837 dictyosteli

34	110.5	2.4	1246	1	VP03_HSVSA	001000 herpesvirus
35	108	2.3	4351	1	PAT2_RAT	088277 rattus norv
36	107.5	2.3	422	1	PAFA_CHICK	Q90677 b platelate
37	106.5	2.3	444	1	PAFA_BOVIN	Q28017 b platelate
38	105	2.3	1442	1	VJ9F_YEAST	P47169 saccharomyc
39	104.5	2.2	673	1	KPCB_BOVIN	P05126 bos taurus
40	104.5	2.2	722	1	V174_RICPR	P81171 rickettsia
41	104	2.2	994	1	BGAL_STAXY	O33815 staphylococ
42	102	2.2	3144	1	VP13_YEAST	O07878 saccharomyc
43	101.5	2.2	664	1	PD13_SHEEP	O02849 ovis aries
44	101	2.2	550	1	INVI_HANAN	P40912 hansenula a
45	100.5	2.2	635	1	TRG4_ECOLI	Q00185 escherichia

ALIGNMENTS

RESULT 1
DPP4_BOVIN STANDARD: PRT: 765 AA.
AC P81425:Q8WMC8;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (adenosine deaminase complexing protein) (ADCP-I)
DE (Activation molecule 3) (ACT3) (WC10).
GN DPP4 OR CD26.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
OC NCBI_TaxID=9913;
CX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Lymphocytes;
RC PubMed=12073152;
RX Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J., Bonnach G.A.;
RA "Molecular characterization of bovine CD26 upregulated by a staphylococcal superantigen.";
RL Immunogenetics 54:216-220(2002).
[2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC TISSUE=Thymus;
RX PubMed=11981836;
RA Gliddon D.R., Howard C.J.;
RT "CD26 is expressed on a restricted subpopulation of dendritic cells in vivo.";
RL Eur. J. Immunol. 32:1472-1481(2002).
[3]
RN SEQUENCE OF 1-24.
RP TISSUE=T-cell;
RC PubMed=11598101;
RX Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K., Naessens J., Bonnach G.A.;
RA "Identity of activation molecule 3 on superantigen-stimulated bovine cells is CD26.";
RL Infect. Immun. 69:7190-7193(2001).
[4]
RN SEQUENCE OF 537-546.
RP TISSUE=Kidney;
RC MEDLINE=98293306; PubMed=9629661;
RX Ben-Shooshan I., Parola A.H.;
RA "The CP-I subunit of adenosine deaminase complexing protein from calf kidney is identical to human, mouse, and rat dipeptidyl peptidase IV.";
RL Comp. Biochem. Physiol. 119B:289-292(1998).
-I- FUNCTION: Removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline. Binds and regulates the activity of ADA.
-I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-I-

RP SEQUENCE OF 281-302.
RC TISSUE=Kidney;
RX MEDLINE=94128239; PubMed=7905271;
RA Iwaki-Egawa S., Matanabe Y., Fujimoto Y.;
RT "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
RT dipeptidyl peptidase IV.";
RL Biochem. Hoppe-Seyler 374:973-975(1993).
CC -1- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC PENULTIMATE RESIDUE IS PROLINE.
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC A SOLUBLE FORM.
CC -1- PTM: THE SOLUBLE FORM (SDP) DERIVES FROM THE MEMBRANE FORM (MDP)
CC BY PROTEOLYTIC PROCESSING.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: J04581; AAA41096.1; -;
CC EMBL: J02997; AAA41272.1; -;
CC PIR: A33315; A33315.
CC DR MEROPS: S09.003; -;
CC DR InterPro: IPR002469; DPPIV_N term.
CC DR InterPro: IPR001375; Peptidase_S9.
CC DR InterPro: IPR002471; Prol_endopep_ser.
CC DR InterPro: IPR000379; Ser_estrs_site.
CC DR Pfam: PF00326; Peptidase_S9; 1.
CC DR Pfam: PF00930; DPPIV_N term; 1.
CC DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
CC DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
CC KM Transmembrane; Glycoprotein; Signal-anchor.
CC KM CHAIN 1 767
CC FT CHAIN 37 767
CC FT CHAIN 281 767
CC FT TRANSMEM 7 28
CC FT DOMAIN 1 6
CC FT ACT_SITE 631 631
CC FT ACT_SITE 709 709
CC FT ACT_SITE 741 741
CC FT CARBOHYD 83 83
CC FT CARBOHYD 90 90
CC FT CARBOHYD 148 148
CC FT CARBOHYD 217 217
CC FT CARBOHYD 227 227
CC FT CARBOHYD 319 319
CC FT CARBOHYD 521 521
CC FT CARBOHYD 666 666
CC FT CONFLICT 38 38
CC FT CONFLICT 183 183
CC FT CONFLICT 332 332
CC FT CONFLICT 352 352
CC FT CONFLICT 394 394
CC FT CONFLICT 562 562
CC FT CONFLICT 624 624
CC SQ SEQUENCE 767 AA: 88003 MW: AA17C0EC6F0E4652 CRC64;

Query Match 10.7%; Score 498.5; DB 1; Length 767;
Best Local Similarity 24.1%; Pred. No. 1.3e-29;

Matches 210; Conservative 121; Mismatches 328; Indels 213; Gaps 38;
QY 63 DESGPHSHRLY-----YLGMYPGSRNSLLY-SEIPKPKVKREALLLSWKOMLDHPQATPH 117
Db 31 DEADAASARLYTLADLYLKNTRFVKYSYSLRWVSQSEYLKQENNTLL-----FNA--E 80
QY 118 HGVSREELLRRKRRLGVRGITSYDFH-SEGLPFLOASNSLPHCRDQKNGKNGFVSPMK 176
Db 81 HG-----NSSIFLENSFFELFGDSISDYSPPDLFVLENYVKOMRHSHTASTSYIDLN 136
QY 177 PLEIKTQSGPRMDPKICAPDAF-PFJNNSDLVANIETGERRLTFCHQGLSNVLD 235
Db 137 KRQLTIEEKIPNTQMTWISQEGHKLAYWKNDIYKIEPHLSHRIT--STKEKVIEN 194
QY 236 PKAGAVATYIOIE-FDRFTGYWCPTASWEGSEGLKTLRILEYEDESVEYIH--VP 291
Db 195 ---GINDWYEEIEFGASALWMSNGAF-----LAAVQPNDCVPLIEYSFYS 240
QY 292 SPALERKTDYSYPRPGSKPKIALKAEFOTDSQKIVSTQEKELQPFSSLPKVEY 351
Db 241 DESLOYPKTYWIIPYKAGAVNPV--KFTLVNTDSLSSTTTTPMOITAP-ASVTTGDHY 297
QY 352 IARAGTRDQKAYAMAFLDPRQOQLVLLPPLFTIPSTENEBQRLASAAVBRNQPYY 411
Db 298 LCDVAVWSEDRIS-----LQWL-----RRIQNTS 321
QY 412 V-----YEEVTNWINVHDIFTFPPOSSEDELCLFLANCKTGFGHLKVTAVLKSOGY 466
Db 322 VMAICDYDKTLTW-----NCPYTOEH-----IETSAV 349
QY 467 DWSEPPSPGDEFR-----CPKEE-----IATSGWEVLAIR 500
Db 350 GMCGRPRAPPHFTSSGSSFYKIVSDKDKYKHCQFOKDKKPPQVCTFTTKGMEVIS-- 407
QY 501 GSKIMWNEETKLVYFOGT-KDPLRHLVYVSEAGAEIVRLTTPGFSHSCSMQSFDMF 559
Db 408 ---LEALTSQLYLYISNEKEMPGGNLVKIQ-----LTDHTNKKC-LSCDLNFB 453
QY 560 VSHYSSVTPPCVHYKLS--GP-----DDPLHKQPRFMASME 597
Db 454 RCQYVSYSLSKEKAYVOLGCRGPLYTLHRSTQKELRVLDDNSALDK-----MQ 506
QY 598 AASCPRDYPPELTFHHTSRDVL-YGMITYKPRALDQKKHPTVLVYGGPOVQVLNNSS 656
Db 507 DVO-----MPSKLDFTVLNETRFWTOMLPRH-FDKSKYPLIDVYAGPSCQADAAE 560
QY 657 KGIKYLRNLN---TLASIGYAVVY-IDGRGSCQGRLEFGALKNOMQOVEDIEDVEGL-QF 711
Db 561 -----RLMWATYLASTENITIVASFDGRSGCYGDKIMAHINKRLGLTEVEDQIEAARG 614
QY 712 VAEKYGFIDLSRAVHIGWSTYGGFLSLMGLIHKPOVRKVALTAGAPVTVMAYDTGYERYW 771
Db 615 L--KMFQVDSKRAVLAIGWSTYGGVYTSVNLGSGSGVFCGIAVAPVSRMEYDDVYTERYW 672
QY 772 DV--PENNOHGVRAGSVALVHEKLPNEPNRLLIHGFIDENYVFFHTNPLVSLIAGRP 829
Db 673 GLETPEDNDHYRNSIVMSRAENE--KQVEYLLIHGTADNDVHFQSOAISKALVDAGVD 730
QY 830 YQLOIYPRNHRHSIRCPESGEHEVTLHLPLQE 861
Db 731 FOAMWTYDDEHGIAASTAHQHIYSHMSHPLQO 762

RESULT 3
DPP4_HUMAN STANDARD: PRT: 766 AA.
ID DPP4_HUMAN
AC P27487:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (TPI03) (Adenosine deaminase complexing protein-2)
DE (ADABP).

GN DPP4 OR ADCP2 OR CD26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NC NCB1_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92329551; PubMed=1352704;
 RA Misumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of human dipeptidyl
 RL peptidase IV, a serine proteinase on the cell surface.";
 RL Biochim. Biophys. Acta 1131:333-336(1992).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95012454; PubMed=7927537;
 RA Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
 RT "Genomic organization, exact localization, and tissue expression of
 RL the human CD26 (dipeptidyl peptidase IV) gene.";
 RN Immunogenetics 40:331-338(1994).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Peripheral blood;
 RX MEDLINE=92325476; PubMed=1352530;
 RA Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
 RT Dahlberg H.N., Schlossman S.F., Morimoto C.;
 RL "Cloning and functional expression of the T cell activation antigen
 RT CD26.";
 RL J. Immunol. 149:481-486(1992).
 RN (4)
 RP ERRATUM.
 RX MEDLINE=93171637; PubMed=8094732;
 RA Tanaka T.;
 RL J. Immunol. 150:2090-2090(1993).
 RN (5)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RA Strausberg R.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN (6)
 RP SEQUENCE OF 1-551 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=92165847; PubMed=1347043;
 RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
 RT Trotot P., Barbat A.;
 RL "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
 RT colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
 RT human coding sequence and changes of dipeptidyl peptidase IV mRNA
 RL levels during cell differentiation.";
 RL J. Biol. Chem. 267:4824-4833(1992).
 RN (7)
 RP SEQUENCE OF 545-766 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=91024044; PubMed=1977364;
 RA Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
 RT "Isolation of a cDNA probe for the human intestinal
 RT dipeptidylpeptidase IV and assignment of the gene locus DPP4 to
 RL chromosome 2.";
 RL Ann. Hum. Genet. 54:191-197(1990).
 RN (8)
 RP SEQUENCE OF 1-31 FROM N.A.
 RX MEDLINE=96067599; PubMed=7487939;
 RA Boehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
 RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific
 RT regulation from a TATA-less GC-rich sequence characteristic of a
 RL housekeeping gene promoter.";
 RL Biochem. J. 311:835-843(1995).
 RN (9)
 RP PARTIAL SEQUENCE.
 RC TISSUE=Kidney;
 RX MEDLINE=93210468; PubMed=8096237;
 RA Morrison M.E., Vijayasardhi S., Engelstein D., Albino A.P.,

RA Houghton A.N.;
 RT "A marker for neoplastic progression of human melanocytes is a cell
 RL surface ectopeptidase.";
 RL J. Exp. Med. 177:1135-1143(1993).
 CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
 CC polypeptides having unsubstituted N-termini provided that the
 CC penultimate residue is proline. Plays a role in T cell activation.
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-1-
 CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
 CC Xcc is neither Pro nor hydroxyproline.
 CC -1- SUBUNIT: Homodimer or heterodimer with Seprase (FAP).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form.
 CC -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
 CC by proteolytic processing.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9.
 CC -1- DATABASE: NAME=PRO; NOTE=CD guide CD26 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd26.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U13735; AAB60646.1; -;
 DR EMBL: U13710; AAB60646.1; JOINED.
 DR EMBL: U13711; AAB60646.1; JOINED.
 DR EMBL: U13712; AAB60646.1; JOINED.
 DR EMBL: U13713; AAB60646.1; JOINED.
 DR EMBL: U13714; AAB60646.1; JOINED.
 DR EMBL: U13715; AAB60646.1; JOINED.
 DR EMBL: U13716; AAB60646.1; JOINED.
 DR EMBL: U13717; AAB60646.1; JOINED.
 DR EMBL: U13718; AAB60646.1; JOINED.
 DR EMBL: U13719; AAB60646.1; JOINED.
 DR EMBL: U13720; AAB60646.1; JOINED.
 DR EMBL: U13721; AAB60646.1; JOINED.
 DR EMBL: U13722; AAB60646.1; JOINED.
 DR EMBL: U13723; AAB60646.1; JOINED.
 DR EMBL: U13724; AAB60646.1; JOINED.
 DR EMBL: U13725; AAB60646.1; JOINED.
 DR EMBL: U13726; AAB60646.1; JOINED.
 DR EMBL: U13727; AAB60646.1; JOINED.
 DR EMBL: U13728; AAB60646.1; JOINED.
 DR EMBL: U13729; AAB60646.1; JOINED.
 DR EMBL: U13730; AAB60646.1; JOINED.
 DR EMBL: U13731; AAB60646.1; JOINED.
 DR EMBL: U13732; AAB60646.1; JOINED.
 DR EMBL: U13733; AAB60646.1; JOINED.
 DR EMBL: U13734; AAB60646.1; JOINED.
 DR EMBL: M74777; AAB60646.1; -;
 DR EMBL: BC013329; AAH13329.1; -;
 DR EMBL: BC013329; AAH13329.1; -;
 DR EMBL: BC013329; AAH13329.1; -;
 DR EMBL: X60708; CA443118.1; -;
 DR EMBL: S79876; AAB35614.1; -;
 DR PIR: S24313; CDH026.
 DR MEROPS: S09.003; -;
 DR Genew: HGNC:3009; DPP4.
 DR MIM: 102720; -;
 DR InterPro: IPR002469; DPP4.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; Prol_endopep_ser.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR Pfam: PF00930; DPP4; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 DR HydroLase: Aminopeptidase; Dipeptidase; Serine protease;
 KW Transmembrane; Glycoprotein; Signal-anchor.
 FT DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
 FT CHAIN 1 766 (MDPP).

```

FT CHAIN 39 766 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT 1 6 (SDPP)
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL)
FT TRANSMEM 7 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT 29 766 (POTENTIAL)
FT ACT_SITE 630 766 EXTRACELLULAR (POTENTIAL)
FT ACT_SITE 708 766 CHARGE RELAY SYSTEM (BY SIMILARITY)
FT ACT_SITE 740 740 CHARGE RELAY SYSTEM (BY SIMILARITY)
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CONFLICT 6 6 K -> R (IN REF. 6)
FT CONFLICT 7 7 V -> I (IN REF. 1)
FT CONFLICT 437 437 S -> I (IN REF. 1)
FT CONFLICT 557 557 T -> I (IN REF. 7)
FT CONFLICT 663 663 D -> E (IN REF. 7)
SQ SEQUENCE 766 AA; 88278 MM; 5FBA42C6662D6117 CRC64;

```

Query Match 10.5%; Score 489.5; DB 1; Length 766;

Best Local Similarity 22.4%; Pred. No. 6e-29; Indels 241; Gaps 35;

Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;

```

QY 47 LIVNAPDFQVOKTDESGPHSHRLYYLGMFGYSGRNSLAYSELPKKYRKALLLSMK 106
DB 26 VLKNG-----TDATADSRKTYTL-----TDLKMTYRIKLKLSLR- 62
QY 107 QMLDHFQATPHHGYVSREELL-----REKRLGVFGITSDYH-SESGFLF 153
DB 63 -----LSDHELYLKQENHLLVENAEYGNSSVFLNENSTDFEGHSHINDYSISPGQPTL 115
QY 154 QASNLSPICRGKNGKGFVSPKPLEIKTQCSGPRMDKICADAPF-FSFTINNSDIWA 212
DB 116 LEYNVKKMRHSYTAHYDNLNKKQLITEERIPNNTQWVSPVGHKLAYWNNDIYK 175
QY 213 NIETGERRLTFCHQGLSNVLDPPKAGVATFVIOE-FDRETYGWCPTASMESEGLK 271
DB 176 IEPNLPYRITW--TGKEDIYN--GITDMVYEEVFSASALAMSPNGF----- 222
QY 272 TLRILEEVDSEVYH---VSPALEERKTDSTRYRTGSKNPKIALKAEFOTDSOG 328
DB 223 ---LAVAGFNDTEVPLIEYSFYSDESLOYPKIVRVYPRKAGAVNPTV--KFEVNTDSIS 277
QY 329 KIVSTQEKELVQFSSLPFKVEYIARAGWTRDGKAYAMFLDRPQOMIQLVLLPPLAFLP 388
DB 278 SVTNAITSQIATPASPMLGD-HYLCDDVWATQERIS-----LQWL----- 316
QY 369 STENEORLASARAVRVNVPYV---YEVEVTWVWVNDIEYFPQSEGEDELCLFLR 443
DB 317 -----RRIONSVMDICDDYDESSGRW----- 337
QY 444 ANECKTGFCHLYKYTAVALKSCGYDMSPEPSEDE-----FK 480
DB 338 --NCLVAROHLEMTST-----GWVGRFRPEPHTLDGNSFYKIIISNEGYRHCYFQ 388
QY 481 CPIKEIALTSGEWEVILARHGSKIWVNETKLIVYFGTKDPLLEHLVYVSEAG----- 536
DB 389 IDKRCCTFTTKGTWVIGLEAL-----TSD-----YLYISNEYGMROG 428
QY 537 -EIVRLTTPGFSSHSCSMQNFDMVSHYSVSTPCCVHYKL--SGP----- 580
DB 429 RNLYKIQDSTYKVCCLCENLPERCQYYSVFSKKAAYKYLQRCGPGPLITLHSSVND 488
QY 561 -----DDDLPHKQPRRWASMEAAASCPDVPVPEIFHFHFRSDVRL-YGAIYYPHALQ 632
DB 489 KGLRVLENSALDK-----MLQNVQ-----MPSKKLDFILNETKFWYQIMLPFH-FD 535

```

```

QY 633 PGKHPTLVFYGGPOVOLVNNSEFKIKYLRN---TLASLGAVV-IDGRSGCORGLR 688
DB 536 KSKKPTLLDDYAGCSQKADTV-----RLNATYIASHTENITVASFDFGSGSYQGD 589
QY 689 FEGALKNGQGYEIEDQVEGL-QFVAERYGFTDLSRVALHGSYGGFTSLMGLHKPOVF 747
DB 590 IMHAINRRLGFEVEDQLEAARQF--SKMGFVDNKRRIAMGWSYGYVTSMVLSGSGVFE 647
QY 748 KVAINGAVTYWMAVDTCYTERYMDV--PENNQHYEAGSVALHKEKLPENRLLIHLG 805
DB 648 KGIIVAVPVSREYEDSVYTERYKMLPPEDNLDHYRNSTVSRANF--KQVEYLLIHLG 705
QY 806 FLDEWVHFHTFVTSOLIRAGKPYOLOIYPERHSIRCPESGEHYETLLHFLQE 861
DB 706 TADDVHQQSQSQISKALVDVGVDFQAMWYDDEDDGASSTAHQIYTHMSHFIRQ 761

RESULT 4
ID DP4_FELCA STANDARD: PRF: 765 AA.
AC Q9N2I7, FELCA
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26).
GN DP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
CX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Periphereal blood;
RX MEDLINE=20094000; Pubmed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue.";
RL Immunogenetics 50:366-368(1999).
CC -I- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline (By similarity).
CC -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-I-
CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC -I- SUBUNIT: Homodimer (By similarity).
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -I- PM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB023952; BAA92344.1; -
CC MEROPS: S09.003; -
CC InterPro: IPR002469; DPPIV_N.term.
CC InterPro: IPR001375; Peptidase_S9.
CC InterPro: IPR002471; ProL_endopep_ser.
CC InterPro: IPR00379; Ser_estrs_site.
CC Pfam: PF00930; DPPIV_N.term.1.
CC Pfam: PF00326; Peptidase_S9.1.
CC PROSITE: PS00708; PRO_ENDOPEP_SER.1.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```

```

FT CHAIN 38 765 (MDPP).
FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT (SDPP) (BY SIMILARITY).
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 7 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 629 629 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 765 AA; 88213 MW; 3EFC98A228175D9 CRC64;

```

```

Query Match: 10.58; Score 486; DB 1; Length 765;
Best Local Similarity 23.08; Pred. No. 1,1e-28;
Matches 200; Conservative 135; Mismatches 331; Indels 204; Gaps 33;

```

```

59 VQKIDSGPHSHRLYLGLMPYSGRENSLLSEIPKVKREALLLSMKQMLDHFQAPPH 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28 LNKQNDAAADSRKRYTL-----TDYLNKTRFVYSLRW-----VSHD 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
119 GVSREBELLRERRRLGVGITSYDFHSEGLFQANSLSLPHCRDGGKNGFVSP--- 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 YLYKQDNNILFLNAEYG-----NSSIFL---ENSTFDEFHSHINDYSVSDQGF 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 -----MKPLEIKQCCGPRNDPRICADPAF--FSFINSLDM 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
113 ILLEYNYVQWRHSYASYDIYDLNKRQLTEERKIPNNNTQITWSPBCHKLAYWKNVY 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
211 VANITGEERLRTCHOGLSNVLDPKSAGAVTAVIOBE--PDRFTGYWCPTASMEGSEG 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
173 VKNEPNSSHRITW--TGEENAIYN---GIADVYEEIEISATSLMWSKGF----- 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
270 LKTLRIIYEEVDESEVEYH--VSPALERKTDSTYRPTGSKNPKIALKLAEPQDTS 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
222 -----LAVAFQNDQVPLDIYSFYSDSLQYPMTRIPYKAGANPTV--KLEVIKTDN 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327 QGKIVSQEOKELVQPFSSLPFKVEYIARAGTRDGKAYAMAFLDLRPOQMLDVLPLALF 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
275 LNPMTNATVEITPP--AAMLTGDDYLLCDVTWANEERIS-----LQMLRIQNYSYMD 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
387 IPSTENEORLASARAVPRNQPVVYEEVNTVMINVDIFYPPQSGDEGLCFLRANE 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
326 IROVNNSTGKRWISSAA-----QEHII--EMSTTGVNGRRRAEPHTSSGRN---FYKIIS 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
447 CKTGECHLYKVTAVLKSGQYDWSPEFSGDEFEKPIKEETALTSGEMEVL----- 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
376 NEDDYKHICR-----FOIDKKDCFTIKGAEVIGIALTTDY 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
498 -----ARHSGKIWNNEETKLYVQGTQKDRPLEHNLVVSF--EAAGEIYALT 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
414 LYYISNEYKMGPGGRNLYKIQLDNYTVVACL--SGELKPERCQYVSFSKSAKKYQLCS 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
543 TPGRS-----HSCSMQNDMFVSHYSSVTPPCVHYVKKLSGDDDLPHKORFNASMMEA 598
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
473 GPGILPLTLHRSSNDE--ELRVLEDNSALDKMLQEVOMPSKKDLPIILNEKRYQMI-- 528
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
599 ASCPPDVVPEPEIFHFRSDVRLXGMTYKPRHALQPKKHPVLEFVGQPOVLYVNNSEFKG 658
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
529 -----LPP---HFDI-----SKKYPILLIYVAGPQCKADALF-- 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
659 IKYIRLN---TLASLGAVVY--IDGRSGCORGLFEGALKMQMQVEEDQVEGL-QFVA 713
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 559 ----RLNMATYLASTENITIVASFDRSGSGYGGKIMAVANRLCTFEVEDQIEARQF-- 612
Qy 714 EKYGFIDLSRAVIAHGWSYGFSLMGLNKRQVFRKVALAGAPVYVMAYDGTERTYMDV 773
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 613 SKMGFVDDKRAIAIAGWSYGGVTSVNLVLAGSGVRCGIAVAPVSRWEYDSYTERIYGL 672
Qy 774 --PENNOHGVAAGVALAHEKLPNEPNRLILHGFLENDEHFFHTNPLVSQILRAGRPYQ 831
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 673 PTPQNDLDYKNSIVMSRAENF--KQVEYLLIHGTADNVHFFQSAQISKALVAGVDFQ 730
Qy 832 LQIYPRNHSIRCPESGEHFEVTLHPLE 861
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 AMWYTDHGDGASGPAHQHIVTHMSHFQK 760

```

```

RESULT 5
DEP4_MOUSE STANDARD; PRT; 760 AA.
AC P28843;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (thymocyte-activating molecule) (THAM).
GN DPP4 OR CD26.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Thymus;
RX MEDLINE=92129288; Pubmed=1370813;
RA Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
RA Pierres M.;
RT "cDNA cloning for mouse thymocyte-activating molecule. A
RT multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
RT subgroup of serine proteases."
RT J. Biol. Chem. 267:2200-2208(1992).
RN [2]
RP REVISIONS.
RA Marguet D.A.;
RA Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.A; TISSUE=Liver;
RX MEDLINE=95092780; Pubmed=7999781;
RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene."
RT Biochemistry 33:15204-15214(1994).
RN [4]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-20.
RX MEDLINE=91302787; Pubmed=1712807;
RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
RA Bernard A.-M., Goyvel J.-P., Pierres M.;
RT "Evidence that thymocyte-activating molecule is mouse CD26
RT (dipeptidyl peptidase IV)."
RT J. Immunol. 147:447-454(1991).
RN [6]
RT FUNCTION: Removes N-terminal dipeptides sequentially from
RN polypeptides having unsubstituted N-termini provided that the
RN penultimate residue is proline.
CC CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-1-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC SUBUNIT: Homodimer.
CC SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form.
CC PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing.

```

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X58384; CAA41274.1; -;
 DR EMBL: U12620; AAA82213.1; -;
 DR EMBL: U12599; AAA82213.1; JOINED.
 DR EMBL: U12600; AAA82213.1; JOINED.
 DR EMBL: U12601; AAA82213.1; JOINED.
 DR EMBL: U12602; AAA82213.1; JOINED.
 DR EMBL: U12603; AAA82213.1; JOINED.
 DR EMBL: U12604; AAA82213.1; JOINED.
 DR EMBL: U12605; AAA82213.1; JOINED.
 DR EMBL: U12606; AAA82213.1; JOINED.
 DR EMBL: U12607; AAA82213.1; JOINED.
 DR EMBL: U12608; AAA82213.1; JOINED.
 DR EMBL: U12609; AAA82213.1; JOINED.
 DR EMBL: U12610; AAA82213.1; JOINED.
 DR EMBL: U12611; AAA82213.1; JOINED.
 DR EMBL: U12612; AAA82213.1; JOINED.
 DR EMBL: U12613; AAA82213.1; JOINED.
 DR EMBL: U12614; AAA82213.1; JOINED.
 DR EMBL: U12615; AAA82213.1; JOINED.
 DR EMBL: U12616; AAA82213.1; JOINED.
 DR EMBL: U12617; AAA82213.1; JOINED.
 DR EMBL: U12618; AAA82213.1; JOINED.
 DR EMBL: U12619; AAA82213.1; JOINED.
 DR EMBL: BC022183; AAH22183.1; -;
 DR PIR: S23752; S23752.
 DR MEROPS: S09.003; -;
 DR MGD: MGI:94919; DPP4.
 DR InterPro: IPR002469; DPP4.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; Prol_endopep_ser.
 DR InterPro: IPR00379; Ser_estrs_site.
 DR Pfam: PF00336; Peptidase_S9; 1.
 DR Pfam: PF00930; DPP4_N_term; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
 KW Transmembrane; Glycoprotein; Signal anchor.
 FT CHAIN 1 760
 FT DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
 FT (MDP).
 FT CHAIN 37 760
 FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
 FT (SDP).
 FT DOMAIN 1 6
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSHEM 7 28
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT CYS-RICH.
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 624 624
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 702 702
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 83 83
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 90 90
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 113 113
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 213 213
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 313 313
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 328 328
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 514 514
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 679 679
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 760 AA: A5F644BA6EA43DB8 CRC64;
 Query Match 10.4%; Score 481; DB 1; Length 760;
 Best Local Similarity 24.6%; Pred. NO. 2.6e-28;
 Matches 214; Conservative 113; Mismatches 326; Indels 218; Gaps 39;

QY 63 DESGSHRLTY-----YLCMPYSGRENSLLY-SEIPKRYKRALLLSWKQMLDHQATPH 117
 DQ 31 DEAAADSRRTSLADLYLKSTFRVKSYSLSMWSDREYLAKQNNILL-----LNAE 80
 QY 118 HGVSREBELLRERKRLGFCITSVDHSES-GLFLFOASNSLPHCDGKNGFVWSPM 175
 DQ 81 HG-----NSIFLENSTFESFG-----YHSVSPDRFLVLEENYKQWMSHTASYANIDV 131
 QY 176 KPLEIKTQCSPPRMDPKICPADPAF-BSFINSDLMVANTF-----TGEERLTF 224
 DQ 132 NKROLITEEKIPNNTQWTWSPBEGHKLAVVKNIDYV-KVPHLPHSRITSTGE----- 185
 QY 225 CHGSLNVLDPKSAGVATFIOEE-FDRPTGYWMCPSAEGSGRLTLRLIYEYDES 283
 DQ 186 -----NVIYN-----GLTDWYEEVEFGAYSALMWSPPNNTF-----LATAQFMDT 225
 QY 284 EVEVIH---VPSPALERKTDYRPTGSKNPKIALALAEFOITDSQKIVSTOEKELVQ 340
 DQ 226 GVPLIEYSFYDESLOQYKTWIPYKAGAVNPYKFFIVMIDISS----- 272
 QY 341 PFSSLPKVEYIARAGWTRDGKY---AMAMFLDRPOQMLVLPLPALFTSPTENEOR 396
 DQ 273 ---SSAAPFIQIPAPASVARGDHYLCDVWATEERISLOML----- 310
 QY 397 LASARAVRNQOPYVY-----YEEV-----TNWVNHDFYRPPSEG 435
 DQ 311 -----RRIQNTSVMAICDYDKINTLWNCPSQOHEMSTGTGVRPAEPFTSDG 362
 QY 436 EDELFLRANECKTGF---CHLYKYTAVLKSQGYDMSPEFSGEDEFKPIKEITALTSQ 492
 DQ 363 SS---FKIISDKDGKYNKICHPK-----DKKCTP-----ITKG 394
 QY 493 EWEVLARHGSKIWNNEETKLYFOGT-KDPLLEHLLYVSYEAGAEIVRLTTPGSHSCS 551
 DQ 395 AMEVIIS-----TEALTSDIYLYISNOYKEMPGKRLYKIQLTDTHTNWKCL-----SCD 442
 QY 552 MSQ-----NPMWFVSHYSSVSPPCVHYKLSGPPDDDLHOPRMAWMMEAASC-- 601
 DQ 443 LNPERCQYAVSFSKEAKYIYQLCMGPGPLPYTLHRSTD---HKELR---VLEDNSALD 495
 QY 602 ---PPDYVPEIFHFRHSRDVL-YGMITYKPHALOPGKHPYLVFYVGGPOVOLVNNFSK 657
 DQ 496 RMLDQVQPKSKKLDPEIVLNETRPFWMQMLPRH-EDKSKKYLLLDYVAGPSCQADASF- 553
 QY 658 GIKYIRLN---TLASLGAIVV-IDRGSCQGRGLFEGALNQMGOVEITEQVBL-QFV 712
 DQ 554 -----RLMWATYLLASTENIIVASEFGSGSGYOGKIMHAINRRLGLEVEDEQIEARQFV 608
 QY 713 AEKYGTDLISVVAIHGWSYGGFLSLMGLIHKPOYKVAIAGAPTYVMAAYDTGTERYM- 771
 DQ 609 --KMGFVDSKRVAILMGWSYGGIVTSMVIGSGSGYFKCGIAVAPVSRMWEYDYSVTERYM 666
 QY 772 -DVENNNGHGEAGSVLHVLEKLPENRRLILHGFLENDEYHFFHTNLSQLIRACKPY 830
 DQ 667 LPREDNDLHNRNSTVMSRAEHF--KQVEYLLIHGTADNVHFFQSSAQISALVADAGVDF 724
 QY 831 QLOIYPNERSHIRCESEGEHYEVLHFLQE 861
 DQ 725 QAMWYTTDEHGIASTAHQIHYSMSHFLQQ 755
 RESULT 6
 ID DPP6_RAT STANDARD; PRT; 859 AA.
 AC P46101;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dipeptidyl peptidase IV like protein (dipeptidyl aminopeptidase-
 DE related protein) (Dipeptidylpeptidase VI) (DPPX).
 OS DPP6.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

[illegible]

Db	289	-ELLKSIHIAHMSPDGT-----	RAYAIINDSRPLMLPLTYTSGSYPTV---	334
Qy	300	TDSYRVPRTGSKNPKIALKIAFQTSQGLKIVSTQEKELVOPFSSLPKVE--YIARGW	357	
Db	335	--PYHPRKSGSENSISLHV-----GLNGPTHLEMPRPD--PRRREYITVWKV	382	
Qy	358	TRDGKVIAMAMLDLRPOOMLOLVLLPRLPFIPTSTENEOGLASARAVRNQPVVYEVT	417	
Db	383	ATSTKVA-VTMIAKQVNSILTLIDATGTCVTKHEE-----	420	
Qy	418	NWMI---NVHDIKFPPOSEGEDELCEFLRANECKTGCHLYKTVAVLKSQGYDMSPPSP	474	
Db	421	EAMLHRQNEEVF-----SKGRKFFRYRAIP-QGGKGRFHNTY-----SSQPNSS	467	
Qy	475	GEDEFKCPKEIEIALTSGSEVLAHRGSKIWNBEETLVYFPGTKDTPLEHNLVVSVEA	534	
Db	468	NDN-----IQ--SITSGDMVT---ELIYDEKRRKKLVFLSTEDPRRHLYTSAN---	512	
Qy	535	AGEIYRLTTPFSHSC-----SMQNFMPVSHSYSTPP-CVN-----	573	
Db	513	-----TVDFNQCISCDLVENCYTVSASFSHNMDFELCKCEGPGVPVTVYHNTDKR	565	
Qy	574	-VYKLSGPDDELKHPREFMAHMEASCPEDYVPELFIHPTBSQVRLGMKYPHALQ	632	
Db	566	RMFDLEANE---QVQAIYDRQPKI---EYKIEV-----EDISLPMQLKPRFTT	611	
Qy	633	PGKKHPTVLFYVGGPOVOLVNSFKGIKYRLNT--LASLGVAVVVIDGRSGCQGLRFE	690	
Db	612	DTAHYPLLLVYDGRFGSGVSERE---VYWEVLVSHGAVVVKDGRSGSGQGRKLL	667	
Qy	691	GALKNQMGQVEIEDQVGLQFAVKYGFIDLSRAVHGMSTYGFSLMGLHK---PQV	746	
Db	668	HEVRRRLRLGFELEKQDMEAVRTML-KEQYIDKTRVAVGKDYGLSTYILPAKENGQGT	726	
Qy	747	EKVAIAGAPVYVMAAYDGYTERMDVPENNQHCEKGSVALVHEKLPNEPNRLLIHGF	806	
Db	727	FTCGSALSPITDFEKLIVASFESERYLGLHGDNRAYETMKLAHRVASL-EDQGLIHAT	784	
Qy	807	LDEWVHFHTFNLVSOLIRACKPQLOIYPNERASINCPESGEHVEYTLHLFDE	861	
Db	785	ADEKHQHTMELLTOLLKGGANSLSQIYDESHYFHSVALKQHLNYSITGFVE	839	
RESULT 7				
DP6P6_HUMAN				
ID	DP6P6_HUMAN	STANDARD;	PRT;	865 AA.
AC	P42658:			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Dipeptidyl peptidase IV like protein (Dipeptidyl aminopeptidase-			
DE	related protein) (Dipeptidylpeptidase VI) (DPPX).			
GN	DPP6.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Hippocampus;			
RA	MEDLINE=93372805; Pubmed=8103397;			
RX	Yocotani N., Dol K., Wenthold R.J., Wada K.;			
RT	"Non-conservation of a catalytic residue in a dipeptidyl			
RT	aminopeptidase IV-related protein encoded by a gene on human			
RT	chromosome 7."			
RL	Hum. Mol. Genet. 2:1037-1039(1993).			
CC	-1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN			
CC	ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE			
CC	FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.			
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).			
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DPPx-L (SHOWN HERE) AND DPPx-S;			

ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M96859; AAA35760.1; -
 CC EMBL: M96860; AAA35761.1; -
 CC MEROPS; S09.973; -
 CC GeneW; HGNC:3010; DPP6.
 CC MIM; 126141; -
 CC InterPro: IPR002469; DPPIV_N.term.
 CC InterPro: IPR001375; Peptidase_S9.
 CC InterPro: IPR000379; Ser_estrs_site.
 CC Pfam; PF00326; Peptidase_S9; 1.
 CC Pfam; PF00930; DPPIV_N.term; 1.
 CC Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 CC DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC DOMAIN 117 865 EXTRACELLULAR (POTENTIAL).
 CC CARBOHYD 173 173 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHYD 319 319 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHYD 404 404 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHYD 471 471 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHYD 535 535 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHYD 566 566 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHYD 813 813 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC VARSPLIC 1 81 MASLYQRTGKINISRPAPPEASHLLGGGPEPDGAGGA
 CC KPLGRQAQAAAPRERGCGGAGGAPRPOYOGRSQGDGE
 CC -> MTTAKEPSASGSKVQOQDE (IN ISOBORN DPPX-
 CC S).
 SO SEQUENCE 865 AA; 97588 MW; 1AB1AE0024464B CRC64;
 Query Match 9.6%; Score 445.5; DB 1; Length 865;
 Best Local Similarity 23.8%; Pred. No. 1.5e-25;
 Matches 176; Conservative 120; Mismatches 268; Indels 175; Gaps 31;

DB 535 NOTYASASHSHMD---FFLAKCEGPVPMVTWNTTDDKKKFFDETNNHKKAIANDROM 591
 QY 608 PEIHFHTR-SDVRLYGMITYKPHALQPGKHPTVLFVYGGPOVOLVNSFKIKYLRLT 666
 DB 592 PKVEYRDLEIDIDYNNPMQILKPATFTDTHYPLLVGDTPSQSGVAEKEF-VSMETV-M 649
 QY 667 LASLGYAVVVDGRSCQGRGLFEGALKNQOVLEIQVDELQVAEKYGTIDISRAVI 726
 DB 650 VSSHAAVAVVCKDGRSGFGTKLLEHVRRLGLLEKQDMENVKML-KEQYIDRTVAIV 708
 QY 727 HWSYGFSLSLGLIHK-----POVEKVAIAGAPVTWMAVDTGYERYWDVPENNQGYE 782
 DB 709 FCKDYGGYSTIYLPAKGNOCOTFTCGALSPIITDFKIYASFSERYIGLGLDNRAVE 768
 DB 769 MKVAVHVSAL--EEQQLIHPTADEKIHQHTAELLITQLIRKANVSLQIYDPESHVF 826
 QY 843 RCPESGEHVEYLLHFLQF 861
 DB 827 TSSSLKQHLNRSIINFVE 845
 RESULT 8
 DPP6_BOVIN
 ID DPP6_BOVIN STANDARD; PRT; 863 AA.
 AC P42659;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dipeptidyl peptidase IV like protein (Dipeptidyl aminopeptidase-
 DE related protein) (Dipeptidylpeptidase VI) (DPPX).
 OS Dipep.
 GN Bos taurus (Bovine)..
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID:9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wade K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: DPPX-L (SHOWN HERE) AND DPPX-S;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
 CC BRAIN, KIDNEY, OVARY AND TESTIS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M76428; AAC41622.1; -
 CC EMBL: M76429; AAC41623.1; -
 CC MEROPS; S09.973; -
 CC InterPro: IPR002469; DPPIV_N.term.
 CC InterPro: IPR001375; Peptidase_S9.
 CC InterPro: IPR000379; Ser_estrs_site.

FT	ACT SITE	734	734	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD	49	49	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	92	92	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	99	99	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	227	227	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	314	314	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	679	679	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	31	35	MISSING (IN ISOFORM 2).
FT	VARSPLIC	31	63	MISSING (IN ISOFORM 3).
FT	CONFLICT	737	737	S -> L (IN REF. 2).
SO	SEQUENCE	761 AA;	87944 MW;	917403AEDAA213B25 CRC64;

Query Match	9.48;	Score 438.5;	DB 1;	Length 761;
Best Local Similarity	22.28;	Pred. No. 4.1e-25;		
Matches 183;	Conservative 126;	Mismatches 287;	Indels 229;	Gaps 35

0Y	118	HGVYSREBELLRERKRRLGVFTISYDF-----HESGELFLQASNSLPHCHDG	160
Db	79	YNIEFRREYITILSNSTKMSVANTDGLSPDQFOYILSDYSKIMRSTAYITAYITDQNG	133
0Y	166	GRNGEWSPMKPLEIKTQSGPRMDPKIC-----PADPAFFSPTNSDLWAN	213
Db	139	--EFVRGVELPRPIQYLCWSP-VGSKLAYUYONNITLQKRGDPPE-----QI	183
0Y	214	IETGERRLTCHOGLSNVLDPPKSAGYATVYIOEEDRFITGY-WMCP7LMSGBGLK	277
Db	184	TYTGRENRI-----FNGIPDWYEBEEM-LATKYALWMSPPDKF-----	220
0Y	272	TLRIIEVEUDESEVEVI-----HYVSPALBEKRTDSTYRPTGNSPKRIALKLAEQ	322
Db	221	---LATVEFNDSIDITLAYSYGQYF-----RTINIPFKAGANNPVRY-----	266
0Y	324	TDSOCKIVSTOEKELVOPFSSLFPKVEIYIAGWTRDGVAMAMEFLDRQ--OMLOLVL	380
Db	265	---FIDTTPYHHVHGEME--VPVPEMIA-----SDYFFSMLTWVSSERVCLOMLKRVQ	311
0Y	381	LRPALFISTENEDEBRLASARAVPRNOPYVEEYVWYINMIDIFRPFQSGEDELIC	444
Db	314	NVSYVSTICDFREDWH---AMECPKN-OEHV--EBSRTGMAGG--FVVS7PAPSQDATS	367
0Y	441	FLRANECKTCFCHLYKATVAVLKSQGYDMSFSPCEDEKCPRIKEIMLTGEMVELARH	500
Db	364	YKIKESDMDGKXHHIYT-----KOTVEALOIDTSGKNEAT---39	
0Y	501	GSKITWNEETKLVY---FOGKTDPLEHNILYVVSYEAAGETVRLTTPGFSHCSM---	555
Db	399	--YIRVYQDSLFYSSNEFEFG--YPRGRNITRILSIGNSPSKCVLT---CHLKER	444
0Y	553	---SQNDMEVSHSSVSTPRCVNYVL-SGPDDDLH-----KOPR--59	
Db	448	QOYVAYSSYAKAYVALCVYGRGRLSTPLDGRKTQEOVLVEENKELENSLRNIOLPRVE	507
0Y	591	---FWMASMEAAASCPRDVPRPEIFHFHTRSDVRLYGMUKYRHALOPKRNPTV	644
Db	508	IKKLKDGGLTWYKMI-----LPPQ-----FDRSKKYPLL	53
0Y	641	LFVIGGPOVOLVNNSE--KGIKYUFLNLTLASIYAUVVIDGRSGQSGRLREGALKOMG	699
Db	538	IQOYVGPSCOSVKSVEFAVNMVITYL---ASKEGIVIALVDRGTAFOGDKFLHAYRKLG	59
0Y	699	QVEIEDOYEGLOFVAKEKGFDLSNVALHSGSYGGFSLMCLINKPOYFKALAGARTV	755
Db	594	VEVEDDOLTAARKFLE-MGFIDEERIALWMSYGGYVSLSLASGTGLFKGCIADAPSS	655
0Y	759	WMAVDTGTERMDPV--ENNNOHGEAGSVALLHVEKLENEENRLLILHGFIDENVFEPT	813
Db	653	MEVYASISYERMGILTPDDNLEHNKNTVAKARAYFRFNV--YLIHG7ADNVHPPNS	713
0Y	817	NFLVSLIRACKPVOLQITPNERHSIRCPSEGENYEVTLLHFLQOE	861
Db	711	AQIAKALVNAQYDFQAMWYSQDNHISGSRQSNHILYTHMTGTEFKQ	755

RESULT	10
DAP2_YEAST	ID DAP2_YEAST STANDARD; PRT; 818 AA.
AC	P18962;
DT	01-NOV-1990 (Rel. 16, Created)
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DPAB B) (YSCV).
GN	DAP2 OR YHR028C.
OS	Saccharomyces cerevisiae (Baker's yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CX	NCBI_TaxID=4932;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=89174971; PubMed=2647766;
RX	Robertes C.J., Pohlig G., Rothman J.H., Stevens T.H.;
RA	"Structure, biosynthesis, and localization of dipeptidyl
RT	aminopeptidase B, an integral membrane glycoprotein of the yeast
RT	vacuole."
RL	J. Cell Biol. 108:1363-1373(1989).
RN	[2]
RC	SEQUENCE FROM N.A.
PC	SPRAIN-S288C / AB972;
RX	MEDLINE=94378003; PubMed=8091229;
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA	Du Z., Favellio A., Fulton L., Gatlung S., Geisel C., Kirsten J.,
RA	Kucaba T., Hillier L., Jier M., Johnston L., Langton V.,
RA	Lafrelle P., Louis E.J., Macri C., Marlis E., Meneses S., Mouser L.,
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA	Vincent D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
VA	Vaudin M.;
RT	*Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT	VIII."
RL	Science 265:2077-2082(1994).
CC	-I SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSSOME-LIKE
CC	VACUOLES.
CC	-I SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X15484; CAAG3512.1; -;
DR	EMBL; U10399; AAB68879.1; -;
DR	PIR; A30107; A30107.
DR	PIR; S46780; S46780.
DR	MEROPS; S09_006; -;
DR	SGD; S0001070; DAP2.
DR	InterPro; IPR002469; DPPIV_N.term.
DR	InterPro; IPR001375; Peptidase_S9.
DR	InterPro; IPR002471; Prol_endopep_ser.
DR	InterPro; IPR000379; Ser_estr_site.
DR	Pfam; PF00326; peptidase_S9; 1.
DR	Pfam; PF00930; DPPIV_N.term; 1.
DR	PROSITE; PS00708; PRO_ENDOPEP_SER_1.
RW	Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
XW	Transmembrane; Glycoprotein; Signal-anchor.
FT	DOMAIN 1 29 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 30 45 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT	DOMAIN 46 818 LOMENAL (POTENTIAL).
FT	ACT_SITE 679 679 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 756 756 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 789 789 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 738 738 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 83 83 O -> H (IN REF. 1).
FT CONFLICT 125 125 S -> N (IN REF. 1).
FT CONFLICT 182 188 FEELING -> LRLLET (IN REF. 1).
FT CONFLICT 200 200 D -> N (IN REF. 1).
FT CONFLICT 366 375 TSNVVRNESS -> DEFRKERK (IN REF. 1).
FT CONFLICT 808 818 AKRAFDOQEVK -> OSVLSMGMLTIELIYSSSHRDHKT
FSYLTMYI (IN REF. 1).
SQ SEQUENCE 818 AA: 93404 MW: 318f45045375BD3 CRC64;

Query Match 9.38; Score 432.5; DB 1; Length 818;
Best Local Similarity 25.4%; Pred. No. 1.3e-24;
Matches 164; Conservative 91; Mismatches 262; Indels 129; Gaps 28;

QY 240 GVAATFVIOEE-FDRFGTYWCPTASWEGSEGLKTLRLIYEVDSEVEVIVHPSPALEER 298
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
D 234 GKPRVYEEVEFEDDKAAMWSPTGDTY-----IAFLKIDSEGEFTIPIYVODEK 283
QY 299 ----KTDYRYPRYSKNPKIALKLAEPQDSDOKIVSTOEKELVQFPSSLFPKVE---- 350
D 284 DIYEMSMIKYKSGTNPNAELMWYSMDG-----TSHPRISGNKK 326
QY 351 ----YIARAGTDRDGKAMAMFLDRPOOMLOLVLPALPIPTSENEORLASARAVPRN 406
D 327 DGSLLILEVYVWNGN-VLVKTTDRSSDILTVFLIDT---IAKTSN----- 368
QY 407 VQPVVVEEYTN--VWINVDHIFRPPQSEGEDELCLFRANEC---KTGRCHLYKTAV 460
D 369 ----VWNESSNGMWITHTL-----FIPANEFTDRPHNGT-----VDI 405
QY 461 LKSGYDWSPEF--SPGEDEFKCPKEIEIALTSGEVEVLARHSGKIWNNEETKLIVFOGTR 519
D 406 LPIGGYHNLAYFEMSNSHK-----TLLEGKEVY--NGPLAFISMENRL-YFISTR 455
QY 520 DTPLEHLLVYVSEAGEIYRLTPG----FSHSCSMQNDMFVSHYSVTPPCV--H 573
D 456 KSSIERHVVYIDLRSPNEIIEVTDSEDGYDVFSFGRRFGLLTYKGPVPYOKIYDFH 515
QY 574 VYKLSGPDPPDLHKOPRFMA5MMEA-ASCPDDY-VPEEIF-----HFHTRSADVLYG 623
D 516 SRRAEKDKGNVLGSKLYHEKNEVLTILEDYAVPRKSRRELNLGDEFGKDLVNSTE 575
QY 624 MIYKPHALDGGKKNPTLVLEVYGGQVQLVNNSEK-GIKYLRNLTLASIGYAVVVDGRGS 682
D 576 ILRPDEFETLSHDHYRFFAYGPGNSQOVVKTSGFENEY---VASQLNLIIVVVDGRGT 632
QY 683 CGSGLRREGALKNOMGVLEIDYEGIQFYAEKYG---FIDLSKVAIHGMSYGGFSLMG 739
D 633 GFGGQDRLSLVRRLDGYEARDOIS---AASLYGSLTFVDPKISLFGMSYGGYTLTKT 688
QY 740 LIHK-POVFVYALAGAVTWMAVDTGTYERYMDVPENNOHGYDAGSVALHVEKLRPEPN 798
D 689 LEKDGGHFKGKSVARPTVMRFDSYTERYMTPTPOENDGYESSY--HNTALAAQAN 746
QY 799 RLILHGEIDENVHFHT-NFIVSQILRACKPYQOLIPNERHSIR 843
D 747 RFLIMHGTGDNDVHFQNSLKFLLDLLDNGVENVDVHVFPSDSHSIR 792

RESULT 11
SEPR_HUMAN
ID 012884; 099998; 000199; 090194; PRT; 760 AA.
AC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
DE membrane serine protease) (170-kDa melanoma membrane-bound
DE gelatinase).
GN FAP.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE-Fibroblast;
RX MEDLINE=94261645; PubMed=7911242;
RA Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncali M.P.,
RA Healey J.H., Old L.J., Rettig W.J.;
RT "Molecular cloning of fibroblast activation protein alpha, a member of
RT the serine protease family selectively expressed in stromal
RT fibroblasts of epithelial cancers."
RT Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE-Melanoma;
RX MEDLINE=97388251; PubMed=9247085;
RA Goldstein L.A., Ghersi G., Pineiro-Sanchez M.L., Salamone M., Yeh Y.,
RA Flessate D., Chen W.-T.;
RT "Molecular cloning of seprase: a serine integral membrane protease
RT from human melanoma."
RL Biochim. Biophys. Acta 1361:11-19(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND
RP 511-518.
RC TISSUE-Melanoma;
RX MEDLINE=97218181; PubMed=9065413;
RA Pineiro-Sanchez M.L., Goldstein L.A., Dote J., Howard L., Yeh Y.,
RA Chen W.-T.;
RT "Identification of the 170-kDa melanoma membrane-bound gelatinase
RT (seprase) as a serine integral membrane protease."
RL J. Biol. Chem. 272:7595-7601(1997).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-Melanoma;
RX MEDLINE=20112818; PubMed=10644713;
RA Goldstein L.A., Chen W.-T.;
RT "Identification of an alternatively spliced seprase mRNA that encodes
RT a novel intracellular isoform."
RL J. Biol. Chem. 275:2554-2559(2000).
RN [5]
RP SEQUENCE OF 192-208; 220-240 AND 510-521.
RX MEDLINE=94327249; PubMed=7519584;
RA Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,
RA Garin-Chesa P., Healey J.H., Old L.J.;
RT "Fibroblast activation protein: purification, epitope mapping and
RT induction by growth factors."
RL Int. J. Cancer 58:385-392(1994).
CC -1- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and may contribute to invasiveness in malignant
CC cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell
CC surface lamellipodia, invadopodia and on shed vesicles.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1/1 (shown here) and
CC 2/s/truncated; are produced by alternative splicing. Isoform 1
CC predominates.
CC -1- TISSUE SPECIFICITY: Fibroblast-specific.
CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling
CC during development, tissue repair, and carcinogenesis.
CC -1- PTM: N-glycosylated.
CC -1- PTM: The N-terminus may be blocked.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



```

FT ACT_SITE 707 707 CHANGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 732 AA; 81292 MW; 3842EB353CE78A40 CRC64;
Query Match
Best Local Similarity 4.6%; Score 215.5; DB 1; Length 732;
Matches 92; Conservative 64; Mismatches 152; Indels 63; Gaps 15;
QY 505 WVNBEETKLVYFGTKDTPLEHNLVYSEAGEIVRLTTPGFSHSCS-MSQNFDMFVSHY 563
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 369 WSAQSQRVPEFSAQRS---RQDLFAVDQ-VGYTSLTAGSGGSKLLTIDQLMAQF 424
QY 564 SSVSTPPCVHYKLSGPDPLHKQRFWASMEASCPDY-----VPEIFHHTR 616
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 425 STSLPPLKLGFLPSPAGKE---QSVLWVSLBEAEPIDIMHGRVLPPE-----Q 474
QY 617 SDVLYGMIKYPHNLQPK-----KHPTVLFVYGGPOVLVNSFKGIKYLRLTSLAG 671
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 475 ENVOYAGLDPEAILLQPSPPDKTQVPMVMPHGGP-----HSSFVYAMMLFPAMLCMG 529
QY 672 YAVVVIDGRSGCQGLRFEGALKNMGCVEIEDQVGLQFAEKY---GFTDLSRAVHIG 728
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 530 FAVLLVNRGSGTGFGQDSILSPGNVGHQVQD---VQFAVEQVLDSEHFDASHVLMG 585
QY 729 WSYGGLSLMGLHKPOVEKVAIAGAVTYVMAYDTGTERYMDVPENNOHGEAG--- 784
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 586 GSHGGLSCHLIGQYPTETRYACAVARNPV-INIASMLGST---DIPD--WCVVEAGPEFS 638
QY 785 -----SVALHEKLPENR-----LILHGFLENVHFHTNPLVSLIRAGRPYQ 831
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 639 SDCLPDLISVAEMLDKSPYRIPQVKTPLLMLGQEDRRVPRFKQGMEXYALKTRNVPR 698
QY 832 LQIYPERHSI 842
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 699 LLTPKSTHAL 709
RESULT 15
ACPH_RAT STANDARD; PRT; 732 AA.
ID ACPH_RAT STANDARD; PRT; 732 AA.
AC P13676; P14320; P70479;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acylamino-acid-releasing enzyme (EC 3.4.19.1) (Acyl-peptide hydrolase)
DE (APH) (Acylaminoacyl-peptidase).
GN APEH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=8925359; PubMed=2722805;
RA Kobayashi K., Lin L.-W., Yeadon J.E., Klickstein L.B., Smith J.A.;
RT Cloning and sequence analysis of a rat liver cDNA encoding acyl-
RT peptide hydrolase."
RT J. Biol. Chem. 264:8892-8899(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=89296508; PubMed=2578023;
RA Lin L.-W., Lee F.J.S., Smith J.A.;
RT "Structural organization of the rat acyl-peptide hydrolase gene.";
RT Nucleic Acids Res. 17:4397-4400(1989).
RN [3]
RP FUNCTION: THIS ENZYME CATALYZES THE HYDROLYSIS OF THE AMINO-
RN TERMINAL PEPTIDE BOND OF AN N-ACETYLATED PEPTIDE TO GENERATE
RN AN N-ACETYLATED AA AND A PEPTIDE WITH A FREE AMINO-TERMINUS.
CC IT PREFERENTIALLY CLEAVES OFF AC-ALA, AC-MET AND AC-SER.
CC -1- CATALYTIC ACTIVITY: Acylaminoacyl-peptide + H(2)O = acylamino acid
CC + peptide.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9C.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04733; AAA88506.1; -
DR EMBL: X14915; CA433040.1; ALT-SEO.
DR PIR: S07624; S07624.
DR MEROPS: S09.004; -.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; Prol_endopep_ser.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
DR HydroLase: Acetylation.
FT MOD_RES 1 1 ACETYLATION (POTENTIAL).
FT MOD_RES 118 118 CHEMICAL NATURE NOT DETERMINED.
FT MOD_RES 291 291 CHEMICAL NATURE NOT DETERMINED.
FT MOD_RES 443 443 CHEMICAL NATURE NOT DETERMINED.
FT ACT_SITE 587 587 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 675 675 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 625 628 IPDM -> M (IN REF. 2).
SQ SEQUENCE 732 AA; 81383 MW; 43F234879E10235B CRC64;
Query Match
Best Local Similarity 4.6%; Score 212.5; DB 1; Length 732;
Matches 87; Conservative 64; Mismatches 162; Indels 53; Gaps 12;
QY 505 WVNBEETKLVYFGTKDTPLEHNLVYSEAGEIVRLTTPGFSHSCS-MSQNFDMFVSHY 563
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 369 WSAQSQRVPEFSAQRS---RQDLFAVDQ-TGTSITSLTAGSGGSKLLTIDQLMAQF 424
QY 564 SSVSTPPCVHYKLSGPDPLHKQRFWASMEASCPDY-----VPEIFHHTR 616
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 425 STSLPPLKLGFLPSPAGKE---QSVSWVSLBEAEPIDIMHGRVLPPE-----Q 474
QY 617 SDVLYGMIKYPHNLQPK-----KHPTVLFVYGGPOVLVNSFKGIKYLRLTSLAG 671
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 475 ENVOYAGLDPEAILLQPSNPDKTQVPMVMPHGGP-----HSSFVYAMMLFPAMLCMG 529
QY 672 YAVVVIDGRSGCQGLRFEGALKNMGCVEIEDQVGLQFAEKY---GFTDLSRAVHIG 728
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 530 FAVLLVNRGSGTGFGQDSILSPGNVGHQVQD---VQFAVEQVLDSEHFDARVALMG 585
QY 729 WSYGGLSLMGLHKPOVEKVAIAGAVTYVMAYDTGTERYMDVPENNOHGEAG--- 776
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 586 GSHGGLSCHLIGQYPTETRYACAVARNPV-INIASMGSTDIPDMCNVETGPPYNSCLPDL 645
QY 777 NOHGYEAGVALHEKLPENRRLILHGFLENVHFHTNPLVSLIRAGKPYOLOIYP 836
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 646 NV--WEEMLDKSPKIKYIPQVKTPLLMLGQEDRRVPRFKQGMEXYALKANVPRLILYP 703
QY 837 NERHSI 842
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 704 KSNHAL 709
Search completed: December 12, 2002, 12:05:01
Job time : 20 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 12:02:02 ; Search time 31 Seconds
(without alignments)
5736.086 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646

Sequence: 1 MATGTGPTADRGDAATDDP.....CPESGHEVETLLHFLQEYL 863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4646	100.0	863	4 Q8WXD8	Q8WXD8 homo sapien
2	2887.5	62.2	883	11 Q9D4G6	Q9D4G6 mus musculus
3	2870	61.8	882	4 Q9HBM5	Q9HBM5 homo sapien
4	2572	55.4	508	4 Q75273	Q75273 homo sapien
5	2383	51.3	439	4 Q9BVR3	Q9BVR3 homo sapien
6	2137.5	46.0	632	4 Q96JX1	Q96JX1 homo sapien
7	1859.5	40.0	432	4 Q75868	Q75868 homo sapien
8	1675	36.1	312	4 Q96NF8	Q96NF8 homo sapien
9	1628.5	35.1	1042	5 Q9VC20	Q9VC20 drosophila
10	1628.5	35.1	1102	5 Q9VC19	Q9VC19 drosophila
11	1494.5	32.2	465	4 Q9HBM3	Q9HBM3 homo sapien
12	1376	29.6	469	4 Q9NXP4	Q9NXP4 homo sapien
13	1220.5	26.3	360	4 Q9HBM2	Q9HBM2 homo sapien
14	1093.5	23.5	310	4 Q9HBM4	Q9HBM4 homo sapien
15	943.5	20.3	927	5 Q96K3	Q96K3 caenorhabdi
16	934.5	20.1	931	5 Q44987	Q44987 caenorhabdi

ALIGNMENTS

RESULT 1

ID	Q8WXD8	PRELIMINARY;	PRT;	863 AA.
AC	Q8WXD8;			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Dipeptidyl peptidase 9.			
GN	PPP9.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Olsen C., Waqtmann N.;			
RT	"Identification and characterization of a novel member of the			
RT	dipeptidyl peptidase IV-related family.";			
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF452102; AAL47179.1;			
DR	InterPro; IPR002469; DPPIV_N_term.			
DR	InterPro; IPR001375; Peptidase_S9.			
DR	InterPro; IPR002410; Pro_annoPTase.			
DR	InterPro; IPR000379; Ser_estrs_site.			
DR	Pfam; PF00930; DPPIV_N_term; 1.			
DR	Pfam; PF00326; Peptidase_S9; 1.			
DR	PRINTS; PR00793; PROAMNOPTASE.			
SQ	SEQUENCE 863 AA; 98263 MW; 40FE0B78E26CDED5 CRC64;			

Query Match 100.0%; Score 4646; DB 4; Length 863;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATGTGPTADRGDAATDDPAAAFQVQKHSNDGLRSIIHSGSKYSGLIVNKAPHDFQFVQ 60

Db 1 MATGTGPTADRGDAATDDPAAAFQVQKHSNDGLRSIIHSGSKYSGLIVNKAPHDFQFVQ 60

Qy 61 KTDESQGHSHRLYLGLMPYSGRENSLLYSEIPKPKYRKALLLSWKQMLDHFQATPHGV 120

|||||

```
Db 61 KTDSGPHSHRLYLGMYPYGSRENSLLYSEIPKVKRKEALLLSWKQMLDHFQATPHGV 120
QY 121 YSREEELLRRKRRLGVGITSYDFHSESGFLFOASNSLHCRDGGKNGFVSPMKPLEI 180
Db 121 YSREEELLRRKRRLGVGITSYDFHSESGFLFOASNSLHCRDGGKNGFVSPMKPLEI 180
QY 181 KTCQSGRMPDKICPADPAFFSFNNSDLVANIETGEERLLTFCHQGLSNVLDPPKSAG 240
Db 181 KTCQSGRMPDKICPADPAFFSFNNSDLVANIETGEERLLTFCHQGLSNVLDPPKSAG 240
QY 241 VAFVIOEEEDRFTGYWMCPTASWEGSEGLKTLRLIYEEVDESEVEIHPVSPALERKT 300
Db 241 VAFVIOEEEDRFTGYWMCPTASWEGSEGLKTLRLIYEEVDESEVEIHPVSPALERKT 300
QY 301 DSVRYPTGSKNPKIALKLAEFQDTSOKIVSTQEKELVQFSSLPKVEYIARAGWTRD 360
Db 301 DSVRYPTGSKNPKIALKLAEFQDTSOKIVSTQEKELVQFSSLPKVEYIARAGWTRD 360
QY 361 GKYAWAMFLDRPOQWLQVLLPALFIPSTENEEQRLASARAVPRNVQYVYVEEVNVW 420
Db 361 GKYAWAMFLDRPOQWLQVLLPALFIPSTENEEQRLASARAVPRNVQYVYVEEVNVW 420
QY 421 INVHDIYPFPQSEGEDELCLFRANECKTGFCILYKVTAVLKSQGYDMSPPSGDEFEK 480
Db 421 INVHDIYPFPQSEGEDELCLFRANECKTGFCILYKVTAVLKSQGYDMSPPSGDEFEK 480
QY 481 CPTKEETALTSGEWEVLARHSGKIWNNEETKLVYFQGTQDTPLEHLLYVVSYEAAGEIVR 540
Db 481 CPTKEETALTSGEWEVLARHSGKIWNNEETKLVYFQGTQDTPLEHLLYVVSYEAAGEIVR 540
QY 541 LTTPGFSHSCSMSONFDMFYSHYSSVSTPPCVHVYKLSGDDDDPLHKQPRFWSMMEAA 600
Db 541 LTTPGFSHSCSMSONFDMFYSHYSSVSTPPCVHVYKLSGDDDDPLHKQPRFWSMMEAA 600
QY 601 CPDYPVPELHFHTRSDVRLYGMIIYKPHALQPKKHPTVLFVYGGPQVQLVNNSEKGIK 660
Db 601 CPDYPVPELHFHTRSDVRLYGMIIYKPHALQPKKHPTVLFVYGGPQVQLVNNSEKGIK 660
QY 661 YLRNLTASLGAVVYIDGRSGCQGLRFEGALKNQMGQVEIEDQVBLQFVAEKYGFID 720
Db 661 YLRNLTASLGAVVYIDGRSGCQGLRFEGALKNQMGQVEIEDQVBLQFVAEKYGFID 720
QY 721 LSRVAIHGWSYGGFLSLMGLIHKPQVKVAIAGAPVTVWMAVDTGYTRYMDVPENNQH 780
Db 721 LSRVAIHGWSYGGFLSLMGLIHKPQVKVAIAGAPVTVWMAVDTGYTRYMDVPENNQH 780
QY 781 YEAGSVALHVEKLPNEPNRLLIHGFLDENVHFFHTNFLVSQILRACKPYQLOIYNR 840
Db 781 YEAGSVALHVEKLPNEPNRLLIHGFLDENVHFFHTNFLVSQILRACKPYQLOIYNR 840
QY 841 SIRCPESGEHYETLLHFLQEYL 863
Db 841 SIRCPESGEHYETLLHFLQEYL 863

RESULT 2
Q9D4G6 PRELIMINARY; PRT; 883 AA.
AC Q9D4G6;
DC 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE 4932434F09RIK protein.
GN 4932434F09RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
```

```
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK016546; BAB30295.1; -.
DR MEROPS; S09_018; -.
DR MGD; MGI:1921638; 4932434F09RIK.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 883 AA; 101297 MW; EA958D6BD33EC71F CRC64;

Query Match 62.2%; Score 2887.5; DB 11; Length 883;
Best Local Similarity 60.5%; Pred. No. 5.8e-223;
Matches 520; Conservative 139; Mismatches 197; Indels 3; Gaps 3;

QY 6 TPTADRCDAATDDPAAR-PQVOKHSHWDGLRSIIHSGRKYSLGLIVNKAPHDFQVQKTDE 64
Db 17 TACEEGNGESQDRKLEPPFYERYSWSQLKLADTRKYHGYMMAKAPHDFMFKRTDP 76
QY 65 SGPHSHRLYLGMYPYGSRENSLLYSEIPKVKRKEALLLSWKQMLDHFQATPHGVYSRE 124
Db 77 DRPHSDRVVYLAANGENRENTLFYSEIPKTNINAAVLMSKPLLDLFQATLDYGMYSRE 136
QY 125 EELLRRKRRLGVGITSYDFHSESGFLFOASNSLHCRDGGKNGFVSPMKPLEIKTQC 184
Db 137 EELLRRKRITGTVIAAYDHPGSGTFLFOAGSGIYHIKDGPHGFTQQPLRPNLVTSC 196
QY 185 SGPRMPKICPADPAFFSFNNSDLVANIETGEERLLTFCHQGLSNVLDPPKSAGVATF 244
Db 197 PNTRMDPKLCPADPDWTFATHSNDIWNLSNLTREERTYVHNELANNEEDPRSGVATF 256
QY 245 VIOEEEDRFTGYWMCPTASWEGSEGLKTLRLIYEEVDESEVEIHPVSPALERKTDYSR 304
Db 257 VLOEEEDRYSGYWCQQAERTPSGG-KILRLIYEEVDESEVEIHPVSPLETRRADSR 315
QY 305 YPTGSKNPKIALKLAEFQDTSOKIVSTQEKELVQFSSLPKVEYIARAGWTRDKYA 364
Db 316 YPKTGTANPKVTFKMSIEIVVDAAGGIDVIDKELVQFPEILFEGVEYIARAGWTPGKHA 375
QY 365 WAMFLDRPOQWLQVLLPALFIPSTENEEQRLASARAVPRNVQYVYVEEVNVNINH 424
Db 376 WSILLDRSQTHLQIVLISPELFIPEDDAMDRORLIESVPDSVTPLIITYEETDINIH 435
QY 425 DIFYFPQSEGEDELCLFRANECKTGFCILYKVTAVLKSQGYDMSPPSGDEFEKCPK 484
Db 436 DIFHVPQTH-EDIEIFIFASECKTGFRLHYKITSILKESKYKRSSGGLPAPSDFKCPIK 494
QY 485 EETALTSGEWEVLARHSGKIWNNEETKLVYFQGTQDTPLEHLLYVVSYEAAGEIVRLTTP 544
Db 495 EETITSGEWEVLGRHSGNIWDEARKLVYFEGTKDSPLEHLLYVTSYANPGEVRLTDR 554
QY 545 GFHSKSCMSONFDMFYSHYSSVSTPPCVHVYKLSGDDDDPLHKQPRFWSMMEAAACPPD 604
Db 555 GYSHSCLSRHCDFFISKYSYQNPKHCVSLYKLSSEDDDPVHKTKFEWATILDSAGPLPD 614
```



```
Qy 605 YVPPIFHFRTRSVRLYGMITYKPHALQPGKKHPTVLFVYGGPQVQLVNNSEFKGIKYLRL 664
Db 615 YTPPEIFSEFTTFTYGLMYKPHDLQPGKKYPTVLFYGGPQVQLVNNRFRKGVKYLRL 674
Qy 665 NTLASLVAVVVIDGRSCQRLRFEKALKNQMGQVELEDQVGLQFVAEKYGDIDLSRV 724
Db 675 NTLASLVAVVVIDNRSGHRLRFEKALKNQMGQVELEDQVGLQFVAEKYGDIDLSRV 734
Qy 725 AIHWSYGGFLSLHGLHKKPQKVAATAGAPVTVMAYDTCTERYMDVPPENNOHGYEAG 784
Db 735 GIHWSYGGFLSLHGLHKKPQKVAATAGAPVTVMAYDTCTERYMDVPPENNOHGYEAG 794
Qy 785 SVALHVEKLPNEPRLHGLHGLHKKPQKVAATAGAPVTVMAYDTCTERYMDVPPENNOHGYEAG 844
Db 795 SVALHVEKLPNEPRLHGLHGLHKKPQKVAATAGAPVTVMAYDTCTERYMDVPPENNOHGYEAG 854
Qy 845 PESGEHVEVLHGLHGLHKKPQKVAATAGAPVTVMAYDTCTERYMDVPPENNOHGYEAG 863
Db 855 PESGEHVEVLHGLHGLHKKPQKVAATAGAPVTVMAYDTCTERYMDVPPENNOHGYEAG 873

RESULT 3
Q9HBM5
ID Q9HBM5 PRELIMINARY; PRT; 882 AA.
AC Q9HBM5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Dipeptidyl peptidase 8.
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221634; AAG29766.1; -.
DR MEROPS; S09.018; -.
DR InterPro; IPR002469; DPPIV_N.term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N.term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 882 AA; 101421 MW; AD801C302DB4652B CRC64;

Query Match 61.88; Score 2870; DB 4; Length 882;
Best Local Similarity 61.5%; Pred. No. 1.5e-221;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;

Qy 24 FOVOKHSDGLRSIITHSGSKYSLGLVKNKAPDFQVOKTDSGPHSHRLYLGLMPYGSRE 83
Db 35 FYEVERYSQKLKLLADTRKHVGYMAKAPDFVKNKNDGPHSDRIYLLANSNGRE 94
Qy 84 NSLYSEIPKVRKREKALLLSSKQMLDFOATPHGHVYSREELLRERKRLGVFGITSYD 143
Db 95 NTLFYSEIPKVINRAVLAALMSKPLDLFQATLDYDGMYSREELLRERKRLGVFGITSYD 154
Qy 144 FHSSEGLFLFOASNLFCRGRGNGFVSPKPLEIKTQSCGPRMDPKICPADPAFTSF 203
Db 155 YHQSGTFLFOAGSGIYHVKDQPGQFTQQLRPLNLVETCPNIRMDPKICPADPAFTSF 214
Qy 204 INNSDLWANTETGERLRTCHOGLSNVLDPPKSGAVATVIOEDRFTGYWMCPTAS 263
Db 215 IHSNDIWSINIVTREERLRTVHNLNNEEDARSAGVATVIOEDRFTGYWMCPTAS 274
Qy 264 WEGSEGLKTLRLYEVEDESEVEIHHVTPSPALEERKTDYRYPRTGSKNPKIALKLPQ 323
Db 274 WEGSEGLKTLRLYEVEDESEVEIHHVTPSPALEERKTDYRYPRTGSKNPKIALKLPQ 323
```

```
Db 275 TTPSGG-KILRLYEDESEVEIHHVTPSPALEERKTDYRYPRTGSKNPKIALKLPQ 333
Qy 324 TDSOGKIVSTOEKELVOPFESLFPKVEIARAGTRDCKYAWAMFLDRPQQLVLVLP 383
Db 334 IDAGRIIDVDELKELIQFELFEGVEIARAGTRDCKYAWAMFLDRPQQLVLVLP 393
Qy 384 ALFIPSTENEBQRLASARAVPRNPQVYVVEEVTNVNINVDIIFYPFQSGEDELCLR 443
Db 394 ELFIPVEDDVMERQLIESVDPSTPLIIEETDIWINIHDIHFVFPQSH-EEIEIF 452
Qy 444 ANEKTGFCHLYKVTAVLKSGYDWSPEFSGEDEFKCPKEETALTSGEVEVLARHGS 503
Db 453 ASECKTGFRLHYKITSILKESYKRSGLPAPSDFKCPKEETALTSGEVEVLARHGS 512
Qy 504 IWNVEKLVYFOGTDPLEHLYVVSVEAAGEIVRLTTPGFHSCSMSONDFMFSHY 563
Db 513 IQDEVRLVYFEGTKDSPLEHLYVVSVEAAGEIVRLTTPGFHSCSMSONDFMFSHY 572
Qy 564 SSVSTPPCVHYKLSGDDDDPLHKQPREWASMEAAASCPDPYVPEIFHFRTRSVRLY 623
Db 573 SNQKNPHCVSLYKLSSPEDDPTCKTEFWATILDSAGPLPDYTPPEIFSEFTTFTY 632
Qy 624 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSEFKGIKYLRLNTLASLGYAVVVIDGRSC 683
Db 633 MLYKPHDLQPGKKYPTVLFYGGPQVQLVNNRFRKGVKYLRLNTLASLGYAVVVIDGRSC 692
Qy 684 QRLRFEKALKNQMGQVELEDQVGLQFVAEKYGDIDLSRVAIHWSYGGFLSLHGLH 743
Db 693 HRGLKFEKALKNQMGQVELEDQVGLQFVAEKYGDIDLSRVAIHWSYGGFLSLHGLH 752
Qy 744 PQVKVAITAGAPVTVMAYDTCTERYMDVPPENNOHGYEAGSVALHVEKLPNEPRL 803
Db 753 SDIFRVAITAGAPVTVMAYDTCTERYMDVPPENNOHGYEAGSVALHVEKLPNEPRL 812
Qy 804 HGFLDENVHFRTHPLVSLQIRACKPYQLIYPNRHSIRPESGEHVEVLHGLHGLH 863
Db 813 HGFLDENVHFRTHPLVSLQIRACKPYQLIYPNRHSIRPESGEHVEVLHGLHGLH 872

RESULT 4
O75273
ID O75273 PRELIMINARY; PRT; 508 AA.
AC O75273;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE R26984_1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lanerudin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Frankheim M., Amico-Keller G.,
RA Cofield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Kobayashi A., Arellano A., Montgomey R., Carrano A.V.,
RA Krommiller B., Arellano A., Montgomey R., Carrano A.V.,
RT "Sequence analysis of a 2.5 Mb region in 19p13.3."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005594; AAC33801.1; -.
DR MEROPS; S09.019; -.
DR InterPro; IPR002410; Pro_annoPase.
DR InterPro; IPR000379; Ser_estrs_site.
DR PRINTS; PR00793; PROAMNOPTASE.
ET NON_TER
SQ SEQUENCE 508 AA; 57750 MW; 2F4D6645BE2D2C89 CRC64;

Query Match 55.4%; Score 2572; DB 4; Length 508;
Best Local Similarity 90.8%; Pred. No. 5.6e-198;
```

Matches 493; Conservative 3; Mismatches 3; Indels 44; Gaps 5;			
Qy	330	IVSTQEKELQVPSSLLPKVEYIARAGWTRDGYAWAMFLDRPOOMQLVLLPALFIPS	389
Db	1	IVSTQEKELQVPSSLLPKVEYIARAG-----AWAMFLDRPOOMQLVLLPALFIPS	53
Qy	390	TENEQRLASARAVPRNVQPVYVVEETVNWVHDIFFYFPQSEGEDELFCFLRANECKT	449
Db	54	TENEQRLASARAVPRNVQPVYVVEETVNWVHDIFFYFPQSEGEDELFCFLRANECKT	113
Qy	450	GFCHLYKVTAVLSQGVDMSEPSGDEEFKCIKEIALTSGEWEVLARHGSKIWNNEE	509
Db	114	GFCHLYKVTAVLSQGVDMSEPSGEG-----EQSLTNA-----IWNNEE	154
Qy	510	TKLVYFGTQDTPLEHHLLYVVSYEAGEIVRLTPGFSHSCSMSONDFVSHYSSVSTP	569
Db	155	TKLVYFGTQDTPLEHHLLYVVSYEAGEIVRLTPGFSHSCSMSONDFVSHYSSVSTP	214
Qy	570	PCVHVYKLSGPDGDDPLHKQPRFNASMMEAAACPPDYVPPFIFHFTSDVRLYGMIIKPH	629
Db	215	PCVHVYKLSGPDGDDPLHKQPRFNASMMEAA-----KIFHFTSDVRLYGMIIKPH	265
Qy	630	ALQPKKHPTVLFYVGGPQVQLVNNNSFGIKYLRNLTLASLGAVVVDGRGSCQGLRF	689
Db	266	ALQPKKHPTVLFYVGGPQVQLVNNNSFGIKYLRNLTLASLGAVVVDGRGSCQGLRF	325
Qy	690	EGALKNQMGVEIEDQVEGLQFVAEKYGFIDLRSVATHGWSYGGFLSLMGLIHKPQVFKV	749
Db	326	EGALKNQMGVEIEDQVEGLQFVAEKYGFIDLRSVATHGWSYGGFLSLMGLIHKPQVFKV	385
Qy	750	AIAGAPVTVMAYDGTGYTERYMDVPENNQHGYEAGSVLHVEKLPNEPNRLLILHGLFDE	809
Db	386	AIAGAPVTVMAYDGTGYTERYMDVPENNQHGYEAGSVLHVEKLPNEPNRLLILHGLFDE	445
Qy	810	NVHFFHTNFLVSLIRACKPYQL-----QIYPNERHSIRCPESGEHYEVTLLHFLQ	860
Db	446	NVHFFHTNFLVSLIRACKPYQLQVALPPVSPQIYPNERHSIRCPESGEHYEVTLLHFLQ	505
Qy	861	EYL 863	
Db	506	EYL 508	
RESULT 5			
Q9BVR3	ID	Q9BVR3 PRELIMINARY; PRT; 439 AA.	
AC	Q9BVR3		
DT	01-JUN-2001	(TrEMBLrel. 17, Created)	
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)	
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)	
DE	Hypothetical 49.9 kDa protein (fragment).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PLACENTA;		
RA	Strausberg R.;		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL:	BC000970; AAH00970.1; -	
DR	MEROPS:	S09.019; -	
DR	InterPro:	IPR001375; Peptidase_S9.	
DR	InterPro:	IPR000379; Ser_estrs_site.	
DR	Pfam:	PF00326; Peptidase_S9; 1.	
KW	Hypothetical protein.		
FT	NON_TER 1		
SQ	SEQUENCE	439 AA; 49926 MW; A18BBA9E12092BAF CRC64;	
Query Match 51.3%; Score 2383; DB 4; Length 439;			
Best Local Similarity 100.0%; Pred. NO. 6.9e-183;			
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			

Qy	425	DIYFPFQSGEGDELCLFLRANECKTGFCCHLYKVTAVLSKQGYDMSPPSPGGEDEFKCPKIK	484
Db	1	DIYFPFQSGEGDELCLFLRANECKTGFCCHLYKVTAVLSKQGYDMSPPSPGGEDEFKCPKIK	60
Qy	485	EEIALTSGEWEVLARHGSKIWNNEETKLVYFGTQDTPLHHLHYVVSYEAAAGEIVRLTTP	544
Db	61	EEIALTSGEWEVLARHGSKIWNNEETKLVYFGTQDTPLHHLHYVVSYEAAAGEIVRLTTP	120
Qy	545	GFSHSCSMSONDFMFVSHYSVSSTPCVHVYKLSGPDGDDPLHKQPRFNASMMEAAASCPDP	604
Db	121	GFSHSCSMSONDFMFVSHYSVSSTPCVHVYKLSGPDGDDPLHKQPRFNASMMEAAASCPDP	180
Qy	605	YVPEPEIHFHFTSDVRLYGMIIKPHALQPKKHPTVLFVYGGPQVOLVNNSPKGIKLYLR	664
Db	181	YVPEPEIHFHFTSDVRLYGMIIKPHALQPKKHPTVLFVYGGPQVOLVNNSPKGIKLYLR	240
Qy	665	NTLASLGAYVVVDGRSCQGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLRSV	724
Db	241	NTLASLGAYVVVDGRSCQGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLRSV	300
Qy	725	AIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTERTYMDVPENNQHGYEAG	784
Db	301	AIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTERTYMDVPENNQHGYEAG	360
Qy	785	SVALHVEKLPNEPNRLLILHGLFDENVHFFHTNFLVSQLIRACKPYQLQIYPNERHSIRC	844
Db	361	SVALHVEKLPNEPNRLLILHGLFDENVHFFHTNFLVSQLIRACKPYQLQIYPNERHSIRC	420
Qy	845	PESGEHYEVTLLHFLQBYL 863	
Db	421	PESGEHYEVTLLHFLQBYL 439	
RESULT 6			
Q96JX1	ID	Q96JX1	PRELIMINARY; PRT; 632 AA.
AC	Q96JX1;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	CDNA FLJ14920 fis, clone PLACE1007416, weakly similar to dipeptidyl		
DE	peptidase IV (EC 3.4.14.5).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PLACENTA;		
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,		
RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yamanoto J., Wakamatsu A., Nakamura Y., Nagaharai K., Masuho Y.,		
RA	Ninomiya K., Iwayanagi T.;		
RT	"NEDO human cDNA sequencing project.";		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK027826; BAB55395.1; -		
DR	MEROPS; S09.018; -		
SQ	SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;		
Query Match 46.0%; Score 2137.5; DB 4; Length 632;			
Best Local Similarity 58.5%; Pred. No. 6.3e-163;			
Matches 395; Conservative 93; Mismatches 134; Indels 53; Gaps			
Qy	189	MDPKICPADPAFFSFINNSDLWANITGTGERLRTFCHQGLSNVLDPPKSGAGVATFVIOE	248
Db	1	MDPKICPADPDWIAFTHSNDIWSINIVTREERRLTYYVHNLANNEDARSAGVATFVIOE	60
Qy	249	EFDRTFGYWCPTASGESEGLTKTLRLIYEVEDESEVEIHPVSPALEERKTSYRPT	308
Db	61	EFDRTFGYWCPTASGESEGLTKTLRLIYEVEDESEVEIHPVSPALEERKTSYRPT	119

Qy	1	MATTGTPTADRGDAADDDPAARFQVQKHSGWDGLRSIIHGSRKYSGLIYNKAPHDFQFVQ	60
Db	72	MATTGTPTADRGDAADDDPAARFQVQKHSGWDGLRSIIHGSRKYSGLIYNKAPHDFQFVQ	131
Qy	61	KTDESGPHSHRLYYLGYMPYGSRENSLLYSEIPKPKVKEALLLLSWKOMLDHFQATPHGV	120
Db	132	KTDESGPHSHRLYYLGYMPYGSRENSLLYSEIPKPKVKEALLLLSWKOMLDHFQATPHGV	191
Qy	121	YSREELLRRKRKLGVEGITSYDFHSGESGLFLFQASNSLFHCRDGGKNGFM-----VS	173
Db	192	YSREELLRRKRKLGVEGITSYDFHSGESGLFLFQASNSLFHCRDGGKNGFMVSPGCCVS	251
Qy	174	PMKPLEIKTQCSGPRMPDKICPADPAFFSFINNSDLWMANIEGTGEERRLTFCHQGLSNVL	233
Db	252	PMKPLEIKTQCSGPRMPDKICPADPAFFSFINNSDLWMANIEGTGEERRLTFCHQGLSNVL	311
Qy	234	DDPKSAGVATFVIOEEDRTGYWMCPTASWEGSEGLKTLRIIYEEVESEVEVIHVPS	293
Db	312	DDPKSAGVATFVIOEEDRTGYWMCPTASWE--EGTKLRIIYEEVESEVEVIHVPS	369
Qy	294	ALEERKTDSDRYPRTGSKNPKIALKLAEFQTDGKIVSTQEKELVQFSSLPKPEYIA	353
Db	370	ALEERKTDSDRYPRTGSKNPKIALKLAEFQTDGKIVSTQEKELVQFSSLPKPEYIA	429
Qy	354	RAG 356	
Db	430	RAG 432	
RESULT 8			
ID	Q96NT8	PRELIMINARY;	PRT; 312 AA.
AC	Q96NT8;		
DT	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)		
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)		
DE	CNNA FLJ30094 fis, clone BNGH4100034, weakly similar to dipeptidyl		
DE	peptidase IV (EC 3.4.14.5).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,		
RA	Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,		
RA	Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,		
RA	Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,		
RA	Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,		
RA	Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,		
RA	Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;		
RT	"NEDO human cDNA sequencing project."		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK054656; BAB70784.1; -		
DR	InterPro; IPR001375; Peptidase_S9		
DR	InterPro; IPR000379; Ser_estr_5ite.		
DR	Pfam; PF00326; Peptidase_S9; 1.		
SQ	SEQUENCE 312 AA; 35518 MW; ABE940AFC5877717 CRC64;		
Query Match 36.1%; Score 1675; DB 4; Length 312;			
Best Local Similarity 99.7%; Pred. No. 3 1e-126;			
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps			
Qy	552	MSQNDFMVSYSVSTPPCVHVYKLSGPDPLHKKOPREWASWMEASCPDYPVPEIF	611
Db	1	MSQNDFMVSYSVSTPPCVHVYKLSGPDPLHKKOPREWASWMEASCPDYPVPEIF	60
Qy	612	HFHTRSDVRLYGLMKPHALQPGKKHPTLVLFVYGGPOVLVNNSFKIKYLRNLTLASLG	671
Db	61	HFHTRSDVRLYGLMKPHALQPGKKHPTLVLFVYGGPOVLVNNSFKIKYLRNLTLASLG	120
Qy	672	YAVVVIDGRSCQRLRFEGALKNQMGQVETEDQVEGLQFVASKYGFIDLSRVAIHGWSY	731

Db	61	HFHTRSDVRLYGMIIYPHALQPCKKHFTVLFVYGGPQVQLVNNSPKGIKYLRLNTLASLG	1200
Qy	672	YAVVVIDGRSCQGLRFEGALKNQMGVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSY	731

```
Db 121 YAVVVYIDGRSCQGRGLRFEGALKNQGVIEDQVEGLQFAVEKYGFDLSRVAIHGWSY 180
QY 732 GGFSLMGLLHKPQVKVATAGAPVTWMAIDYGYTERYMDVPENNQHGYEAGSVLHVE 791
Db 181 GGFSLMGLLHKPQVKVATAGAPVTWMAIDYGYTERYMDVPENNQHGYEAGSVLHVE 240
QY 792 KLPNEPNRLLILHGFGLDENVHFHTNFLVSQILIRAGKPYQLOIYTPNERHSIRCPESGEHY 851
Db 241 KLPNEPNRLLILHGFGLDENVHFHTNFLVSQILIRAGKPYQLOIYTPNERHSIRCPESGEHY 300
QY 852 EYVLLHFLQBYL 863
Db 301 EYVLLHFLQBYL 312

RESULT 9
Q9VC20 PRELIMINARY; PRT: 1042 AA.
AC Q9VC20;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CG3744 protein.
GN CG3744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aquayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A.S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.R., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Poillard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003749; AAF56357.1;
```

```
DR MEROPS; S09.016; -.
DR FlyBase; FBgn0039240; CG3744.
DR InterPro; IPR002469; DPIP_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPIP_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 1042 AA; 116706 MW; 5B915D5C365DF937 CRC64;

Query Match 35.1%; Score 1628.5; DB 5; Length 1042;
Best Local Similarity 39.1%; Pred. No. 1.1e-121;
Matches 370; Conservative 143; Mismatches 315; Indels 119; Gaps 20;

QY 3 TTGTP----TADRGAAA-----TDDPA-----ARFQVKHSDGLRSIIHGRKSYGL 47
Db 124 TSGTPPHGLDVDEGDEDDPVDNDGHIAAPTNPKSWAENKQVQVIRKMCN 183
QY 48 IVNKAHPDFQVOKTDESGPHSHRLYYLGMYPGSRNSLLSYEIPKKVKEALL-----102
Db 184 LSSWVPTNVQFRHLSGDA-----RCYFLGTPPOSWETTLFADINLTQSEEOQLLVORLE 239
QY 103 -----LSWKOMLDH-FOATPHG-----VY 121
Db 240 GIASDEWSPTMNAGSPTSSGHPAFLFNSLPRRLPWSPLLOQPIQSSGGSGSGSASPY 299
QY 122 SREEELLRRKRLGVFGITSYDFHSESGFLFQASNSLHFCRDGKGNFMVSPMKPLEIK 181
Db 300 AREYQLLOERKRLSTWGITSYELHKPSGLVPCFNDLYOCLDTGYNSGL---LFTQLR 356
QY 182 TQSGPRMPKICPADPAFFSFNNSDLMVANITGEERRLTFCHOGLSNVLDDPKSAGV 241
Db 357 TCPQWTALDPQICPNQSDMIAYISDCDLFVTHPLSGHEKRLTYTSGRHSYVDDALSAGV 416
QY 242 APTVIOEEEDRFTGYWMCPTASWEGSEGLKTLRLIYEEVDESEVEVIHVPSPALEKRTD 301
Db 417 PSYVMEERFSRYQGFWMQP-----HSNDGI--YRIVVEVDESEVSYTPSPSTAMGRVD 470
QY 302 SYRYPRTGKNPKIALKLAFAEFTDSQGIKIVSTQEKELVQVFPFSLFPKVEYIARAGTRDG 361
Db 471 EYRFPRTGPNKSKLKLQVLEALQVSEITAKDLPYSLLAFLVSWLEIVYRGTGTPDA 530
QY 362 KYAWAFILDRPQWLQVLVLPALFIPSTENEEQRASA-----RAVPRNVQP-VYVY 413
Db 531 KYVWVGLDRKQOQRLVDILIPDNFCESYSSQVSTPTDSIGDHSWRSLSRTITPLQVY 590
QY 414 EEVTVNWINVHDFYFPQSEGEDELCEFLRANECKTGFCHLYKVTA-VLKSQGYDWSEPF 472
Db 591 TERSDSWINVHDMHLFDLT--ETSVTFLWASB-ETGFRHLVLTASLLLSQANGQPDG 647
QY 473 SPGE-----DEFKCPIKEEIALTSGEWEVLARHGSKIWNBEETKLVIYFGQTKDTPLEH 525
Db 648 SVGAQPSFVDSLALQPRILNKVALTSGEVELARN---LWVDKANKLVYFVGLDTPLEK 704
QY 526 HLYVVSYEAGEIVRLITTFGSHSCSMSONDFVSHYSVSTPVCVHYVKKLGGPDDP- 584
Db 705 HLYVVSLEPERHLLTEPGSYLVLEFDDQCKLMLLVLYCNIQRLPCKVMRNVQTCSSNG 764
QY 585 -----LHKQPRFWASMEAAACPDYVPPPIEFHFHTSDVRLYCYMIYKPHALQP 633
Db 765 VNGIQISLVGYLH-----EGGKPEQYC-POIFSPQLPSGDIVIVAMWFKPHNFEL 813
QY 634 GKHKPTFLVYGGPQVQLVNNFSGIKYLRNLNTLASLAVVYIDGRSCQGRLEFEGAL 693
Db 814 GVKYPTVLNVYGGPEVQVTNNTFKGKHLRMHMLAAQGCVCICIDSGSRHRHGRKRESHI 873
QY 694 KNOMGOVEITDQVEGLQFAVEKYGFDLSRVAIHGWSYGGFLSLMGLIHKPQVKVAIAG 753
Db 874 RGRMGVELTDQVDAIRSLSDQLGYIDMDRVATHGWSYGGYLSGLVGLVYQPKFKVAIAG 933
QY 754 APTVYMAVYDGTGYTERYMDVPENNQHGYEAGSVLHVEKLPNEPNRLLILHGFGLDENVHF 813
Db 934 APTVYMAVYDGTGYTERYMDVPENNQHGYEAGSVLHVEKLPNEPNRLLILHGFGLDENVHF 993
```



```
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Dipeptidyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
DR Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221636; AAG29768.1; -.
DR MEROPS; S09.018; -.
FT NON_PER 1
SQ SEQUENCE 465 AA; 53197 MW; 22939EC0A4FE29CE CRC64;

Query Match 32.2%; Score 1494.5; DB 4; Length 465;
Best Local Similarity 50.5%; Pred. No. 1.8e-111;
Matches 281; Conservative 69; Mismatches 105; Indels 101; Gaps 2;

QY 308 TGSKNPKIALKAEFQDSQKIVSTQEKELVQPFSSLPFKVEYIARAGWTRDGYAWAM 367
DB 1 TGTANPKVTFKMEIMIDABRIIDVIDKELIQPFELFEGVEYIARAGWTPGKYAWSI 60

QY 368 FLDRPQWLQVLVLPALFIPSTENEOQLASARAVPRNPVQVYVEETVNWVHDF 427
DB 61 LLRSQTRQLVILSPFLFVDDVMEQRLIESPDSVTPVLIYEETDWINIHDF 120

QY 428 YPEPQSEGEDELFLANECKTFCFLYKVTAVLKSGQVDSPEPSPGDEKCPKEEI 487
DB 121 HVPQSH-EEIEFIFASECKTGRHLYKITSILKESKYRSGGLPAPSDFCPKPEEI 179

QY 488 ALTSGEWELARGSKIWNEETKLVYFQGTDPLEHLYVYVYEAAGSIVRLTTPGES 547
DB 180 AITSGEWELGRHGSNIQDVRRVYFEGTKDSPLEHLYVYVYVNPGEVRLTRDGY 239

QY 548 HSCSMQNFDMFVSHVSSVTPPCVHVYKLSGDDPLIKQPRFASMEAAACPDYVP 607
DB 240 HSCCISQHCDFTSKYSNQKNPCVSLYKLSPEDDPTCKTKEFNATLDSAGPLDPTP 299

QY 608 PELFHFTRSDVRLYGMVYKPHALQPKKHTVLFVYGGPOVLVNSFKGIKYLRLNTL 667
DB 300 PELFSESTGTGLYGLYKPHDLQPKKHTVLFYGGPO----- 340

QY 668 ASLGYAVVWIDGRGSCQGLRFGALKNQMGVEIEDQVEGLQFVAEKYGFIDLSRVAIH 727
DB 341 ----- 340

QY 728 GWSYGGFLSMLGLHKPQVKVAIAGAPVTVMAYOTGTYTERVMDVPENNHQYEGSVA 787
DB 341 -----VAIAGAPVTLWTFYDTGTYTERVMDVPHDQNEQYGLGSVA 379

QY 788 LHVEKLPNEPRLILHGLFDENVHFFHNFVLSQLIRAGKPYQLQIYDNERHSIRCPES 847
DB 380 MQAEKPEPSEPNRLLHGLFDENVHFFHNFVLSQLIRAGKPYQLQIYDNERHSIRCPES 439

QY 848 GEHYEYTLHLFLQEYL 863
DB 440 GEHYELHLLHYLOENL 455

RESULT 12
Q9NXXF4
ID Q9NXXF4 PRELIMINARY; PRT; 469 AA.
AC Q9NXXF4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE CDNA FLJ20283 fis, clone HEP04088.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
DR Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221636; AAG29768.1; -.
DR MEROPS; S09.018; -.
SQ SEQUENCE 469 AA; 54367 MW; 088EED0B1E46C11F CRC64;

Query Match 29.6%; Score 1376; DB 4; Length 469;
Best Local Similarity 56.1%; Pred. No. 6.1e-102;
Matches 245; Conservative 85; Mismatches 105; Indels 2; Gaps 2;

QY 24 FOVKHSDGLRSIIHSGSRKYLIVNKAPHDFQVQKTDSEGHSHRLYYLGMPTGSR 83
DB 35 FYVERYSWQKLLADTRKYHGYMAKAPHDFMFKRNDPDGPHSDRIYYLAMSNGRE 94

QY 84 NSLLYSEIPKKVRKEALLLSWKQMLDHFQATPHHGVYSREELLRKRLGVFGITSYD 143
DB 95 NTLFYSEIPKTTINRAAVMLSWKPLDLFOATLDYGMYSREELLRKRLGVFGITSYD 154

QY 144 FISESGFLFPOASNSLFLHCRDGGKNGFMVSPMKPLIKTCQCSGRMDPKICPADPAFFSF 203
DB 155 YHOGSGTFLFAGSGGLYHVKGDPGQGTQOPLRPNLVETSCPNIRMDPKICPADPAFF 214

QY 204 INNSDLWANIEGERRLFTFCOGLSNVLDLDPKSAVATFVIOEEDFDRFTGYWMCPTAS 263
DB 215 IHSNDWISNIVTREERLTYVHNLANMEEDARSAGVATFVIOEEDFDRFTGYWMCPTAS 274

QY 264 WEGSEGLTLRLIYEEVDESEVEVHVSPALAEERKTSYRPTGSKNPKIALKLAEFQ 323
DB 275 TTPSGG-KILRLIYEEVDESEVEVHVSPALAEERKTSYRPTGSKNPKIALKLAEFQ 333

QY 324 TDSOGKIVSTQEKELVQPFSSLPFKVEYIARAGWTRDGYAWAMFLDRPQWLQVLVLP 383
DB 334 IDAEGRIIDVIDKELIQPFELFEGVEYIARAGWTPGKYAWSILLDRSOTRQLQVILSP 393

QY 384 ALFIPSTENEOQLASARAVPRNPVQVYVEETVNWVHDFIFPFQSEGEDELCLFLR 443
DB 394 ELFIPVEDDVMERQRLIESVPSVTPVLIYEETDWINIHDFHVFQSH-EEIEFIF 452

QY 444 ANECKTGFCHLYKVTAV 460
DB 453 ASECKTGFCHLYKITSI 469

RESULT 13
Q9HBM2
ID Q9HBM2 PRELIMINARY; PRT; 360 AA.
AC Q9HBM2;
DT 01-WAR-2001 (TrEMBLrel. 16, Created)
DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-WAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Dipeptidyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```



```
SQ SEQUENCE 927 AA; 105365 MW; 6907C6AAEF829D0F CRC64;

Query Match      20.3%; Score 943.5; DB 5; Length 927;
Best Local Similarity 29.4%; Pred. No. 9.6e-67;
Matches 263; Conservative 146; Mismatches 316; Indels 169; Gaps 34;

QY 72 LYILGMPYGSRENSLLYSEIPKVY-----RKEALLLSWKOMLDHFQATPHHGVYSR 123
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 82 MYAIISSVPGTNTQISFTIPLELVEKAQVADRKFELKLSGYNVDSYIRKTPPSAEFTL 141
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 124 EELLREKRKLGIVGITSYDFHSEGLFLFOASNSLFH----- 161
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 142 QCE--RORSQV-VTGISDIYEI--RNGRMILMAGDQLFRYNPLNEALAAPIAVPDDOSST 196
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 162 -----CRDGGKNGFMVSPMKPLEIK---TQCSPRMDP-----KI 193
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 197 EPMDISBGSITSGTKGCSNEAPOSSTVPPVTRIPRIKPTTSTKPTATAPTNNFVSSAKV 256
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 194 CPADPAFFSFINNSDLNWNIEGTERRRLTFCHGLSNVLDDPKSAGVATFVIOEEDRF 253
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 257 CPADSSLLAYVLNKQVYI-----EKNKIIHRTSSN--SKHITNGVPSYIVOEELERF 307
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 254 TGYWMCPTASWEGSEGLTKLRLIYEEVDESEV-----IHWPSPALEERTDSYRPTG 309
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 308 EGIWW-----SES-KT-RLLYEHVNEEKVAESQFGVNGDPEVA-----PMKYPRAG 351
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 310 SKNPKIALKLAETFDSQGIKIVSQEKELVQPFSSLPKVEYIARAGWTRDGKYAWAMFL 369
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 352 TKNAYSTLRWVILE--NGRAYDPLKDEV--IYKHCFFYEYITRAGFFSDGTTVWVQVM 406
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 370 DRPOOILQVLVLLPALPFPSTE-----NEEORLASARAVPR--- 405
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 407 SRDAQCSLLIIPYDFLLPEELGGSIKEDNLQSLTDNLNGVMDKSHETMEKPPRGKL 466
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 406 --NVQPVVYVEEYTNVINVDIFYPQSEGEDELCLFLRANECKTGFCFLYKVTAVLK 462
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 467 RGTVQ---IHKARDYWINTHNAYPLKIDTEHPMYEFYICLEKPNKNGSC-LALISAEILD 522
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 463 SQGYDNSEPFSPGDEFKCPKEIALTSGEWELARHGSKIWNNEETKLVIYFOGTDTP 522
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 523 QNGY-----CRHTEKLLMAENFSINKSMG--IVVDEVRELVYVYVANESHP 566
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 523 LEHLIYVYVEAGEIVRLTTPGFSSHCSMSQN-----FDM----FVSHYSSVSTPPCVH 573
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 567 TEWNI-CVSHYRTGHAQLTESGI--CFKSERANGKLALDLHDHGFACYMTSVGSPAECR 622
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 574 VYKLSGDDDLPHKQPRFASMEASCP--PD--VYVPEIFHFT--RSDVRLYGMLYKP 628
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 623 FYSRWNENEVL--PSTVYAANITVSGHPGQPDJHFDSPEMIEFQSKKTGLMHYAMILRP 680
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 629 HALQPKKKHPTLVFYGGPOVLVNSFKGIKYLRLNTLASLYAVVYVVDGRCSCQRLR 688
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 681 SNFDPYKKYPVFHYVYGGPGIIVHNDFSWQIYR---FCRLGYVVVVIDNRGSAHRGIE 737
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 689 FEGALKNQMGQVETEDQVEGLQFVAEY--GFIDLSRVAIHGWSYGGFSLMGLIHKPOVF 747
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 738 FERHIIKKMGTEVEDQVEGLQMLAERTGGFMDMSRVVHGWSYGGYNALQMIKHPNIY 797
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 748 KVATAGAPVTVMAYDGYTERYDVPDENNOHGYEAGSVALHVEKLPNEPRLILHGLF 807
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 798 RAAJAGGAVSDMRLYDFAITYERNYGP--LEEYVYGASSITGLVEKLPDEPNRLMLVHGLM 856
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 808 DENVHFHTNFLVSQLIRAGKPYQLOIYPNERHSIRCPESGEHYEVTLLHFLOE 861
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 857 DENVHFAHLTHLVDCEIKKGKWHELVIFPNERHGVNRNDASIYLDARMYFAOQ 910
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
```

Search completed: December 12, 2002, 12:05:51
Job time : 35 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 12, 2002, 12:10:08 ; Search time 63 seconds
(without alignments)
4200.985 Million cell updates/sec

Title: US-09-976-674-3
Perfect score: 4646
Sequence: 1 MATTGTTATADGAAATDDP.....CPESGHEHYETLLHFLQEYL 863

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09976674/runat_04122002_162531_7161/app.query.fasta_1.1031
-DB-issued_Patents_NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blos62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09976674.ecgn.1.1.26 @runat_04122002_162531_7161 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	532.5	11.5	823	4	US-09-280-116-171
2	489.5	10.5	2924	5	PCT-US93-07923-1
3	474	10.2	5496	4	US-09-462-284-1
4	426	9.2	2815	1	US-08-230-491A-1
5	426	9.2	2815	1	US-08-619-280A-1
6	426	9.2	2815	2	US-08-940-391-1
c 7	327.5	7.0	4982	3	US-08-699-103B-1
c 8	327.5	7.0	4982	4	US-09-229-059-1
c 9	318	6.8	543	4	US-09-221-017B-253
c 10	299.5	6.4	657	4	US-09-221-017B-646
c 11	267	5.7	612	4	US-09-392-184-31
12	240.5	5.2	1896	3	US-09-016-080-2

c 13	216.5	4.7	3085	4	US-09-221-017B-1010	Sequence 1010, App
14	195.5	4.2	502	4	US-09-280-116-172	Sequence 172, App
15	162.5	3.5	2636	1	US-08-227-689-1	Sequence 1, Appli
16	162.5	3.5	2636	4	US-09-368-169-1	Sequence 1, Appli
17	153	3.3	1869	2	US-08-684-646A-1	Sequence 1, Appli
18	153	3.3	1869	2	US-09-066-285-1	Sequence 1, Appli
19	153	3.3	1869	3	US-09-261-006-1	Sequence 1, Appli
20	153	3.3	1869	3	US-08-951-088-1	Sequence 1, Appli
21	153	3.3	1869	4	US-09-609-566-1	Sequence 1, Appli
22	153	3.3	1869	4	US-09-609-570-1	Sequence 1, Appli
23	132.5	2.9	815	4	US-09-221-017B-962	Sequence 962, App
24	118.5	2.6	6029	4	US-09-385-742B-8	Sequence 8, Appli
25	113.5	2.4	1041	2	US-08-602-359A-24	Sequence 24, Appli
26	111	2.4	1876	1	US-08-483-232-24	Sequence 24, Appli
27	111	2.4	1876	2	US-08-485-938A-24	Sequence 24, Appli
28	111	2.4	1876	2	US-08-910-041-24	Sequence 24, Appli
29	111	2.4	1876	3	US-09-348-474-24	Sequence 24, Appli
30	111	2.4	1876	3	US-09-100-546-24	Sequence 24, Appli
31	111	2.4	1876	3	US-09-010-715-24	Sequence 24, Appli
32	111	2.4	1876	4	US-09-577-758-24	Sequence 24, Appli
33	110.5	2.4	1533	1	US-08-483-232-23	Sequence 23, Appli
34	110.5	2.4	1533	2	US-08-485-938A-23	Sequence 23, Appli
35	110.5	2.4	1533	2	US-08-910-041-23	Sequence 23, Appli
36	110.5	2.4	1533	3	US-09-328-474-23	Sequence 23, Appli
37	110.5	2.4	1533	3	US-09-100-546-23	Sequence 23, Appli
38	110.5	2.4	1533	3	US-09-010-715-23	Sequence 23, Appli
39	110.5	2.4	1533	4	US-09-577-758-23	Sequence 23, Appli
40	106.5	2.3	4220	4	US-09-183-846A-11	Sequence 11, Appli
41	106.5	2.3	4220	4	US-08-961-578C-11	Sequence 11, Appli
42	105.5	2.3	2465	5	PCT-US92-08090-1	Sequence 1, Appli
43	103.5	2.2	4149	2	US-08-737-715-1	Sequence 1, Appli
c 44	103	2.2	2211	4	US-09-318-448-26	Sequence 26, Appli
c 45	103	2.2	2211	4	US-09-347-878-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-280-116-171
; Sequence 171, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280.116A
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 171
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: prolyl oligopeptidases
US-09-280-116-171

Alignment Scores:
Pred. No.: 8.03e-50 Length: 823
Score: 532.50 Matches: 110
Percent Similarity: 76.27% Conservative: 25
Best Local Similarity: 62.15% Mismatches: 28
Query Match: 11.46% Indels: 14
DB: 4 Gaps: 2

US-09-976-674-3 (1-863) x US-09-280-116-171 (1-823)

Qy	699	GlnValGluIleGluAspGlnValGluGly-LeuGlnPheValAlaGluLysTyrGlyPhe	718
Dy	3	CAATAGAAATGACGATCAGGTGGAGGACATCATCTAGCTTCGATATGATT	62
Qy	718	eIfAspLeuSerArgValAlaIleHisGlyTyrSerTyrGlyGlyPheLeuSerLeuMe	738

Qy	329	LysIleValSerThrGlnGluIysGluLeuValGlnProPheSerSerLeuPheProLys	348
Db	842	TCAGTCAACAACTCACTTCCATACAAATCACTGCTCTGCTTCTATGTTCATAGGGGAT	901
Qy	349	ValcIuTyrlleAlaArgAlaGlyTrpThrArgaspGlyLysTyrlaTatPalaMetPhe	368
Db	902	---CACTACTTGTGTGATGTGACATGGGCACACAAAGAAATAATTCF---	946
Qy	369	LeuAspArgProGlnTrpLeuGlnLeuValLeuLeuProAlaLeuPheIlePro	388
Db	947	-----TTGCAGTGGCTC-----	958
Qy	389	SerThrGluAsnGluGlnArgIleuAlaSerAlaArgAlaValProArgAsnValGln	408
Db	959	-----AGGAGGATTCAG-----	970
Qy	409	ProTyrrValVal-----TyrGluGluValThrAsnValTrpIleAsnVal	423
Db	971	AACTATTCCGTCATGGATATTGTGACTATGATGAATCCAGTGGAAAGATGG-----	1021
Qy	424	HisAspIlePheTyrrProPheProGlnSerGluGlyGluAspGluLeuCysPheLeuArg	443
Db	1021	-----	1021
Qy	444	AlaAsnGluCysLysThrGlyPheCysHisLeuTyrlLysValThrAlaValLeuLysSer	463
Db	1022	-----AACTGCTTTAGTGGCAGCGCAACACATGAATGAGTACTACT-----	1063
Qy	464	GlnGlyTyrrAspTrpSerGluProPheSerProGlyGluAspGlu-----	478
Db	1064	-----GGCTGGGTGGAAGATTAGGCTTCAGAACCTCATTTTACCCTTGTATGGT	1114
Qy	479	-----PheLys	480
Db	1115	AATAGCTTCTACAAGATCATCAGCAATGAGAGGTTACACACACATTTGCTATTTCCAA	1174
Qy	481	CysProIleLysGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaAargHis	500
Db	1175	ATAGATAAAAGAACTGCACATTTATTACAAAAGCACCTCGGAAGTCATCGGGATAGAA	1234
Qy	501	GlySerLysIleTrpValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp	520
Db	1235	GCCTCA-----ACCATGAT-----	1249
Qy	521	ThrProLeuGluHisLeuTyrrValValSerTyrrGluAlaAlaGly-----	536
Db	1250	-----TATCTATACTACATTAGTATGAATATAAGGAATCCAGGAGGA	1294
Qy	537	---GluIleValArgLeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsn	555
Db	1295	AGGAATCTTTATAAAATCCAACTTAGTACTATACAAAGTGACATGCCTCAGTTGTGAG	1354
Qy	556	PheAspMetPheValSerHisTyrrSerValSerThrProProCysValHisValTyr	575
Db	1355	CTGAATCCGGAAAGGTGCAGTACTATTCTGTGTCATTTCAGTAAAGAGCGCAAGTATTAT	1414
Qy	576	LysLeu-----SerGlyProAspAspProLeuHis-----	586
Db	1415	CAGCTGATGATGTTCCGGTCTCTGG-TCTGCCCTCTATACTCTACACACACCGCTGAATGA	1473
Qy	587	-----LysGlnPro---ArgPheTrpAlaSerMetMetGluAlaAlaSerCysPro	602
Db	1474	TAAAGGGCTGAGAGTCTCGAAGACAAATTCAGCTTTGGATAAAATGCTCGAGAATGT---	1530
Qy	603	ProAspTyrrValProProGluIlePheHisPheHisThrArgSerAspValArgLeu-Ty	622
Db	1531	CCAGATGCCCTCCAAAAAACTGGACTTCATTTATTTTGAATGAAACAAATAH---TTGGTA	1587
Qy	622	rGlyMetIleTyrrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPhe	642
Db	1588	TCAGATGATCTGCCTCCCTCAT---TTTGATAAATCCAGGAATATCTCTACTTACTAGA	1644
Qy	642	eValTyrrGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrrLe	662

Db	1645	1646	1647	1648	1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661	1662	1663	1664	1665	1666	1667	1668	1669	1670	1671	1672	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682	1683	1684	1685	1686	1687	1688	1689	1690	1691	1692	1693	1694	1695	1696	1697	1698	1699	1700	1701	1702	1703	1704	1705	1706	1707	1708	1709	1710	1711	1712	1713	1714	1715	1716	1717	1718	1719	1720	1721	1722	1723	1724	1725	1726	1727	1728	1729	1730	1731	1732	1733	1734	1735	1736	1737	1738	1739	1740	1741	1742	1743	1744	1745	1746	1747	1748	1749	1750	1751	1752	1753	1754	1755	1756	1757	1758	1759	1760	1761	1762	1763	1764	1765	1766	1767	1768	1769	1770	1771	1772	1773	1774	1775	1776	1777	1778	1779	1780	1781	1782	1783	1784	1785	1786	1787	1788	1789	1790	1791	1792	1793	1794	1795	1796	1797	1798	1799	1800	1801	1802	1803	1804	1805	1806	1807	1808	1809	1810	1811	1812	1813	1814	1815	1816	1817	1818	1819	1820	1821	1822	1823	1824	1825	1826	1827	1828	1829	1830	1831	1832	1833	1834	1835	1836	1837	1838	1839	1840	1841	1842	1843	1844	1845	1846	1847	1848	1849	1850	1851	1852	1853	1854	1855	1856	1857	1858	1859	1860	1861	1862	1863	1864	1865	1866	1867	1868	1869	1870	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052</
----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	--------

Alignment Scores:
Pred. No.: 7.43e-42 Length: 5496

Score: 474.00 Matches: 180
Percent Similarity: 41.34% Conservative: 97
Best Local Similarity: 26.87% Mismatches: 277
Query Match: 10.20% Indels: 117
DB: 4 Gaps: 28

US-09-976-674-3 (1-863) x US-09-462-284-1 (1-5496)

QY 193 IleCysProAlaAspPro-----AlaPhe-PheSerPheIleAsnAsnSerAspLeuTr 210
DB 2365 ATATGCCCAATGGAGCCCGTGGCAATACCATCGCTTTTGTTCGCGAGATGACCTTTA 2424
QY 210 pValAlaAsnIleGluThrGluGluArgArgLeuThrPheCysHisGlnGlyLeuSe 230
DB 2425 TGTC-----TGGGATAATGGTACCGTTACTCGCATTA----- 2458
QY 230 rAsnValLeuAspAsp-----ProLysSerAlaGlyValAlaThrPheValIleGL 247
DB 2459 -----GATGATGGTGGCCCGACATGTTCCACGGCGTGCAGACTGGATCTATGA 2508
QY 247 nGluGluPhe-----AspArgPheThrGlyThrTrpCysProThrAlaSerTrp-- 264
DB 2509 AGAGGAGATCCTCGGCATCGCTAC-----GGCTGTGTGTTCTCGCCGATGGTGAATATCT 2565
QY 265 -----GluGlySerGluGlyLeuLysThrLeuArgIleLeuThrGluGluVa 280
DB 2566 GGCTTACTTGAGCTTCAATGAGACTGGGTTCGACCTACACCGTTTCAGTAT---TATAT 2622
QY 280 lAspGluSerGluValGluValIleHisValProSerProAlaLeuGlu---GluArgLy 299
DB 2623 GGATAACAAGAGATC-----GCTCCGGCGGTATCCATGGGAGCTGAA 2664
QY 299 sThrAspSerTyArgTyProArgThrGlySerLysAsnProLysIleAlaLeuLysLe 319
DB 2665 G-----ATAGGTATCCCAAGGTGTCGACAGCAATCCGACCGTGAAGTCT 2715
QY 319 uAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuVa 339
DB 2716 G-----CTTAACATCGCTAGCAAGGAGGTGAA 2742
QY 339 lGlnProPheSerSerLeuPheProLysValGluTyIleAlaArgAlaGlyTrpThrAr 359
DB 2743 GCAGGCCCGCATCGACCGTTCGAGTCAACTGACTTGCATCTT----- 2785
QY 359 gAspGlyLysTyArgTyAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuV 379
DB 2786 -----GGCAGGTGCTTGGCTCA-----CTGA 2808
QY 379 alLeuLeuProAlaLeuPheIleProSerThrGluAsnGluGluArgLeuAlas 399
DB 2809 TACTCACACCACCGTCTGCTAAGCGGTTCAACCGGTCCAGGACCAGCAAAAGGTCTG 2868
QY 399 erAlaArgAlaValProArgAsnValGlnProTyArgValValTyGluGluValThrAsnV 419
DB 2869 CGCGGTGCATCTGCTCCGCAACAG-GCTACTGTCATCAGCGACCGATGGGACCGATG 2927
QY 419 alTrpIle---AsnValHisAspIlePheTyProPheProGlnSerGluGlyGluAspG 438
DB 2928 GATGGCTCGATAACCTTCTTCAATGAAGTATATTTGGCCCTATCAACCGTCCGCAAGG 2987
QY 438 luLeuCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyTrpLysValT 458
DB 2988 ATGCCCTACTACATGACATCTCTGAC---CATTGGGATGGGCGCATCTGTATCTCTTC- 3043
QY 458 hrAlaValLeuLysSerGlnGlyTyArgTrpSerGluProPheSerProGlyGluAspG 478
DB 3044 -----CCCGTTTCGGCGGC----- 3058
QY 478 luPheLysCysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpGluValLeuA 498
DB 3059 -----GAACCTATCCCATCAACCAAGGCGATGGGAGGTACAG- 3097
QY 498 laArgHisGlySerLysIleTrpValAsnGluGluThrLysLeuValTyPheGlnGlyT 518

DB 3098 -----TCTATTCTGAGTATTGATCAGGAACGCCAGTTGGTGTACTACCTGTGCA 3146
QY 518 hrLysAspThrProLeuGluHisHisLeuTyArgValValSerTyArgGlu-----AlaAlaG 536
DB 3147 CTCAACACACACAGCAGCGGCGCATCTACTCCGCTCTCTATTCACGCTTCGCGTCA 3206
QY 536 lyGluIleValArgLeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsn 556
DB 3207 CCGCGCTCGTCGACACACCGTTCGCCGCTAC---TGGTCTGCTTCTTCTCCCGGACT 3263
QY 556 heAspMetPheValSerHisTySerSerValSerThrProCysValHisValTyTrL 576
DB 3264 CGGGCTACTACATCTCATACATACGAGGCCAGACGTA---CCCTACAGGAACCTCTACA 3320
QY 576 ysLeuSerGlyProAspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetM 596
DB 3321 CGACCAACAGTAC-----AAACCACTCCCGCAATCACCACAAACG 3362
QY 596 etGluAlaAlaSerCysProProAspTyArgValProGluIlePheHisPheHis----- 614
DB 3363 CCAAGTACTCGACCAATCAAGGACTATGCATTTGCCCAACATCCTACTTCCGAGTTC 3422
QY 615 --ThrArgSerAspValArgLeuTyArgMetIleTyLysProHisAlaLeuGlnProG 634
DB 3423 CCGTCCCTCCGGAGAAACCTCAATGTGTGACAGCGTTACCCCGCGGTCTCCCGG 3482
QY 634 lyLysLysHisProThrValLeuPheValTyArgGlyProGlnValGlnLeuValAsnA 654
DB 3483 ATAAGAAGTACCCCATACTTTTACCCCATACGCGCGCCAGCGCCCAAGAGTACCA 3542
QY 654 snSerPheLysGlyIleLysTy---LeuArgLeuAsnThrLeuAlaSerLeuGlyTyAr 673
DB 3543 AGATGGCAAGCCCTGAATTTCAAGGCCTATGTCGCTCCGACAGCAACTCCAGTACG 3602
QY 673 laValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaL 693
DB 3603 TAACCTGAGTGTGACACACCGCGGACAGGTTTCAAGGACGCAAGTTCCGCTCCCGG 3662
QY 693 euLysAsnGlnMetGlyValGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValA 713
DB 3663 TCAGCGCCCAACTCGGCTCTCTCGAAGACAGACCGAGATCTACGCGCGGCA---CAGG 3719
QY 713 laGluLysTyArgGlyPheIleAspLeuSerArgValAlaIleHisGlyTrpSerTyArg 733
DB 3720 CGGCCACATCCCTGGATCGATCGACACCATCGGCATCTGGGCTGGAGTTTCGGAG 3779
QY 733 lyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaG 753
DB 3780 GCTACTTGACAGCAAGGTCTCTGGAGAGGACAGCGGTGCTTTCACATTTAGGAGTCA 3839
QY 753 lyAlaProValThrValTrpMetAlaTyArgThrGlyTyThrGluArgTyMetAspV 773
DB 3840 CCGCCCTGTTTTCGACTGGCGTTTCTACGACTCAATGTACACGAGCGCTACATGAAGA 3899
QY 773 alProGluAsnAsnGlnHisGlyTyArgGluAlaGlySerValAlaLeuHisValGluLysL 793
DB 3900 CCCTCTCGACCAATCAGGAGGCTACGAGCACCGCGCTCCGC---AAGACTGACGGGT 3956
QY 793 euProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnValHisP 813
DB 3957 TCAAGAAGCTCGAGGCGGATTTCTTGATCCAGCACGGAACGCGGCGAGTAACTCCATT 4016
QY 813 hePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyArgGlnLeuG 833
DB 4017 TCCAGAATCCGCTGCGCTGGATCTCTGTGATGGCGATGGC----- 4060
QY 833 lnIleTyProAsnGluArgHisSer 841
DB 4061 --GTCCTCTCTGAGAACCTCCATTTCG 4084
RESULT 4
US-08-230-491A-1

; Sequence 1, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
; TITLE OF INVENTION: ISOLATED OLIGONUCLEOTIDE MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FELFE & LYNCH
; STREET: 805 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT - ASC II
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/230,491A
; FILING DATE: 20-APRIL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587299man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2815 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: linear
; US-08-230-491A-1

Alignment Scores:
Pred. No.: 5.43e-37 Length: 2815
Score: 426.00 Matches: 173
Percent Similarity: 39.08% Conservative: 106
Best Local Similarity: 24.23% Mismatches: 254
Query Match: 9.17% Indels: 182
DB: 1 Gaps: 31

US-09-976-674-3 (1-863) x US-08-230-491A-1 (1-2815)

Qy	195	ProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTrrpValAlaAsnIle	214
Db	734	CCAGGAGATCCACCTTTT-----	751
Qy	215	GluThrGlyGluGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAsp	234
Db	752	-----CAATAACATTT-----AATGAAGAGAAATAATAATTT	787
Qy	235	AspProLysSerAlaGlyValAlaThrPheValIleGlnGluGluPheAspArgPheThr	254
Db	788	AAAT-----GGAATCCAGAGCTGGGTTTATGAAGAGGAATG---CTTCCCTACA	832
Qy	255	GlyTyr-----TrrpTrrpCysProThrAlaSerTrrpGluGlySerGluGlyLeuLysThr	272
Db	833	AAATATGCTCTCTGGTGTCTCTCAATGGAAAAATTT-----	868
Qy	273	LeuArgIleLeuTyrGluGluValAspGluSerGluValGluValIleHisValProSer	292
Db	869	-----ITGGCATATGCGGATTTTAAATGATAAGATATACAGTTATTCCTATTCCTAT	922
Qy	293	ProAlaLeuGluGlu-----ArgLysThrAspSerTyrArgTyrProArgThrGlySerLys	311
Db	923	TATGGCGATGAACAATATCTCTAGAACAAATAATATTCATACCCAAAGGCTGGAGCTAAG	982

Qy	312	AsnProLysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleVal	331
Db	983	AATCCCGTTGTTCCGATA-----TTTATTATC	1009
Qy	332	SerThrGlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysVal-----	349
Db	1010	GATACCACTTACCTCGGTATAGTAGTCCCCAGGAAGTCCCTGTTCAGCAATGATAGCC	1069
Qy	350	-----GluTyrIleAlaAlaGlyTrrpThrArgAspGlyLysTyrAlaTrpAla	366
Db	1070	TCAAGTGATTATTATTCAGTTGGCTCAGGTGGTTACTGATGACGAGTATGT-----	1123
Qy	367	MetPheLeuAspArgProGlnGlnTrrpLeuLeuLeuLeuLeuProProAlaLeuPhe	386
Db	1124	-----TTCCAGTGGCTAAAAAGAGTCCAGAAATGTTTCGGTCTGTCT	1165
Qy	387	Ile-----ProSerThrGluAsn-----	392
Db	1166	ATATGTGACTTCAGGAAGACTGGCAGACATGGGATTGTCCAAAGACCCAGGACATATA	1225
Qy	393	GluGlnArgLeu-----AlaSerAlaArgAlaValProArgAsnValGlnProTyr	410
Db	1226	GAAGAAACAGACACTGGATGGCTGGTGGATTCTTTGTTTCAAGACCCAGTTCAGCTAT	1285
Qy	411	ValValTyrGluGluValThrAsnValTrrpIleAsnValHisAspIlePheTyrProPhe	430
Db	1285	-----	1285
Qy	431	ProGlnSerGluGlyGluAspGluLeuCysPheLeuArgAlaAsnGluCysLysThrGly	450
Db	1386	-----GATGCCATTTTCGTACTACAAAATATTAGTGACAAAGGATGGC	1327
Qy	451	PheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGlu	470
Db	1328	TACAAACATATTCTACTATATC-----	1348
Qy	471	ProPheSerProGlyGluAspGluPheLysCysProIleLysGluGluIleAlaLeuThr	490
Db	1349	-----AAACACACTGGGAAATGCTATTCAAAATTACA	1381
Qy	491	SerGlyGluTrrpGluValLeuAlaArgHisGlySerTyrIleTrrpValAsnGluGluThr	510
Db	1382	AGTGGCAAGTGGGAGGCCATA-----AATATATTACAGAGTAACACAGGAT	1426
Qy	511	LysLeuValTyrPheGlnGlyThr-----LysAspThrProLeuGluHisLeuTyr---	528
Db	1427	TCACTGTTTATTCTAGCAATGAATTTGAAGAATACCTCGGAAGAAAGAACATCTACAGA	1486
Qy	529	-----ValValSerTyrGluAlaAlaGlyGluIleValArgLeuThrThrProGlyPhe	546
Db	1487	ATTAGCATTTGAAGCTATCTCCCAAGCAAGAGTGTGTT-----	1525
Qy	547	SerHisSerCysSerMet-----SerGlnAsnPheAspMet	558
Db	1526	-----ACTTCCCATCTAAGGAAAGAGGTGCCAATATTACACAGCAAGTTTCAGCGAC	1579
Qy	559	PheValSerHisTyrSerSerValSerThrProCysValHisValTyrLysLeu---	577
Db	1580	TACCCCAAGTACTATGACATTTGTCTGTACGGCCAGCATCCCATTCACCCCTTCAT	1639
Qy	578	SerGlyProAspAspAsp-ProLeuHis-----LysGlnProArgPheTrrpAlaSerMe	595
Db	1640	GATGAGCCACTGATCAAGAATAATAATCTCGGAAGAAACACAGGAATTTGAAATGCT	1699
Qy	595	tMetGluAlaAlaSerCysProProAspTyrValProProGluIlePheHisPheHisTh	615
Db	1700	TTGAAAAATATCCAGCTGCCT-AAAGAG-----GAAATTAAGAACTTGAAGT	1746
Qy	615	rArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLys	635
Db	1747	AGATGAATTAATCTTTATGGTACAGATGATCTTCTCTCTCTCAA---TTTGACAGATCAA	1803

```
QY 635 sLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSe 655
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 1804 GAAGTATCCCTTGTCAATGATCAAGTGTATGGTGGTCCCTGCAGCTCAGAGTGTAGGCTGT 1863
QY 655 rPhe-----LysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAl 673
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 1864 ATTTCGTGTTAATGGATATCTTATCTT-----GCAAGTAAGGAAGGATGGT 1911
QY 673 aValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLe 693
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 1912 CATTCGCTTGGTGGTGGTGCAGGAACAGCTTCCAGAGTGACAACTCCTCTATGCGAGT 1971
QY 693 uLysAsnGlnMetClyGlnValGluIleGluAspGlnValGlnGlyLeuGlnPheValAl 713
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 1972 GTATCGAAGCTGGTGTATGAAGTTGAAGACCACTTACAGCTGTGCAGAAATTCAT 2031
QY 713 aGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisClyTyrPheTyrGlyG 733
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2032 AGAA---ATGGGTTTCATTGATGAAAAGAAATAGCCATATGGGCTGGTCTATGGAGG 2088
QY 733 yPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaG 753
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2089 ATAGCTTTCATCACTGCCCTTCGATCGCACTGGCTCTTTTCAATGTGGTATAGCAGT 2148
QY 753 yAlaProValThrValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspVa 773
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2149 GGCTCCAGCTCCAGCTGGGAATATACGCGTCTGTCTACACAGAGAGATTTCATGGGTCT 2208
QY 773 lPro-----GluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValG 791
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2209 CCCAACAAAGGATGATAATCTTGAGCACTATAAGAAATTCACACTGTGATGCGCAAGAGCAG 2268
QY 791 uLysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnVa 811
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2269 ATATTCAGAAATCTAGAC-----TATCTCTCATCCAGGACAGCAGATGATAATGT 2322
QY 811 lHisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGl 831
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2323 GCACCTTCAAAACTCCAGCAGATGCTGTAAGCTCTGGTAAATGCACAACTGGATTCCA 2382
QY 831 nLeuGlnIleTyrProAsnGluArgHisSerIleArgCysProGluSerGly----- 848
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2383 GGCATGTGTACTCTGACCAGAACACCGGCTTA-----TCCGGCTGTCCAC 2430
QY 849 -GluHisTyrGluValThrLeuLeuHisPheLeuGlnGlu 861
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2431 GAACCACTTATACCCACCATGACCCTTCTTAAGACGAG 2470

RESULT 5
US-08-619-280A-1
; Sequence 1, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2815 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-619-280A-1

Alignment Scores:
Pred. No.: 5,43e-37 Length: 2815
Score: 426.00 Matches: 173
Percent Similarity: 39.08% Conservative: 106
Best Local Similarity: 24.23% Mismatches: 254
Query Match: 9.17% Indels: 182
DB: 1 Gaps: 31

US-09-976-674-3 (1-863) x US-08-619-280A-1 (1-2815)

QY 195 ProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIle 214
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 734 CCAGGAGATCCACCTTT-----CAATAACATTT-----AATGGAAGAGAAATAAATATTT 751
QY 215 GluThrGlyGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAsp 234
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 752 -----CAATAACATTT-----AATGGAAGAGAAATAAATATTT 787
QY 235 AspProLysSerAlaGlyValAlaThrPheValIleGlnGluPheAspArgPheThr 254
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 788 AAT-----GGAATCCAGACTGGGTTTATGACAGAGAAATG---CTTCTTACA 832
QY 255 GlyTyr-----TrpTrpCysProThrAlaSerTrpGluGlySerGluGlyLeuLysThr 272
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 833 AAATATGCTCTCTGGTGGTCTCTCTAATGGAATTT----- 868
QY 273 LeuArgIleLeuTyrGluGluValAspGluSerGluValGluValIleHisValProSer 292
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 869 -----TTGGCATATCGGAATTTAATGATAGAGATATACCATGTTATTCCTATTCCTAT 922
QY 293 ProAlaLeuGluGlu---ArgLysThrAspSerTyrArgTyrProArgThrGlySerLys 311
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 923 TATGGCGATGACAAATATCTCTAGACAAATAAATATTCATACCAAGAGCTGGAGCTAAG 982
QY 312 AsnProLysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleVal 331
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 983 AATCCGTTGTTTCGGATA-----TTTATTATC 1009
QY 332 SerThrGlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysVal----- 349
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 1010 GATACCACTTACCCTGGGTATGATGAGTCCAGAGAGTGGCTTCCAGCAATGATAGCC 1069
QY 350 -----GluTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyrAlaTrpAla 366
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 1070 TCAAGTGATTATTATTTCAGTTGGCTCAGCTGGGTTACTGATCAACGAGTATGT----- 1123
QY 367 MetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProAlaLeuPhe 386
|||:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 1124 -----TTGCAGTGGCTAAAAAGAGTCCAGAAATGTTTCGGTCTGTCT 1165
QY 387 Ile-----ProSerThrGluAsn----- 392
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 1166 ATATGTGACTTCAGGAAGACTGGCAGACATGGGANTTGTCCAAAGACCAGGACATATA 1225
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
```

```
Qy 393 GluGluGlnArgLeu-----AlaSerAlaArgAlaValProArgAsnValGlnProTyr 410
Db 1226 GAAAGAAAGCAGAACTGGATGGCTGGTGGATCTCTTTGTTTCAAGACAGATTTTCAGCTAT 1285
Qy 411 ValValTyrGluGluValThrAsnValTrpIleAsnValHisAspIlePheTyrProPhe 430
Db 1285 ----- 1285
Qy 431 ProGlnSerGluGlyGluAspGluLeuCysPheLeuArgAlaAsnGluCysLysThrGly 450
Db 1286 -----GATGCCATTTTCGTACTACAAAATATTATTAGTGACAAGATGGC 1327
Qy 451 PheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGlu 470
Db 1328 TACAAACATATTCACTATATC----- 1348
Qy 471 ProPheSerProGlyGluAspGluPheLysCysProIleLysGluGluIleAlaLeuThr 490
Db 1349 -----AAGACACTGTGGAAAATGCTATTCAAAATTACA 1381
Qy 491 SerGlyGluTrpGluValLeuAlaArgHisGlySerLysIleTrpValAsnGluGluThr 510
Db 1382 AGTGGCAAGTGGAGGCCATA-----AATATATTCAGATGAACACAGAT 1426
Qy 511 LysLeuValTyrPheGlnGlyThr---LysAspThrProLeuGluHisHisLeuTyr--- 528
Db 1427 TCACCTGTTTTATTAGCAATGAATTTGAAGAATACCTCGGAAGAAGAACATCTACAGA 1486
Qy 529 -----ValValSerTyrGluAlaAlaGlyGluIleValArgLeuThrProGlyPhe 546
Db 1487 ATTAGCATTTGGAAGCTATCTCCCAAGCAAGAAAGTGTGT----- 1525
Qy 547 SerHisSerCysSerMet-----SerGlnAsnPheAspMet 558
Db 1526 -----ACTGCCATCTAAGAAAGAGGTGCCAATATTACACAGCAAGTTTCAGCCAC 1579
Qy 559 PheValSerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeu--- 577
Db 1580 TACGCCAAGTACTATGCACCTGTCTGTCTACGCCCGCAGCAGCATCCCATTTCCACCCCTTCAT 1639
Qy 578 SerGlyProAspAspAsp-ProLeuHis-----LysGlnProArgPheTrpAlaSerMe 595
Db 1640 GATGGACGACCTGATCAAGAAATTTAAATCTCGGAAGAAACAAAGGAATTTGGAAATGCT 1699
Qy 595 tMetGluAlaAlaSerCysProProAspTyrValProProGluIlePheHisPheHisTh 615
Db 1700 TTGAAAATATCCAGCTGCCT-AAAGAG-----GAAATTAAGAAACTTGAAGT 1746
Qy 615 rArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLy 635
Db 1747 AGATGAATTTACTTTATGGTACAAGATGATTTCTCTCTCTCAA---TTTGACAGATCAA 1803
Qy 635 sLysHisProThrValLeuPheValTyrGlyProGlnValGlnLeuValAsnSe 655
Db 1804 GAAGTATCCCTGTGTAATTAAGTGTATGGTGGTCCCTGCAGTCAGAGTGAAGTCTGT 1863
Qy 655 rPhe-----LysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAl 673
Db 1864 ATTTGCTGTTAATTTGGATATCTTATCTT-----CCAAGTAAGGAAGGATGGT 1911
Qy 673 aValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLe 693
Db 1912 CATTCCTTGGTGGATGGTCAGGAACAGCTTCCAAGGTGCACAACTCCTCTATGCAAT 1971
Qy 693 uLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAl 713
Db 1972 GTATCGAAAGCTGGGTGTTTATGAAGTTGAAGACAGATACAGCTGTGCAGAAAATTCAT 2031
Qy 713 aGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTrpSerTyrGlyGl 733
Db 2032 AGAA---ATGGGTTTCATTTGTAAGAAAAGAAATAGCCCATATGGGGCTGCTATGGAG 2088
Qy 733 yPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaGl 753
```

```
Db 2089 ATAGCTTTTCATCACTGGCCCTTGCACTCTGGAACCTGGTCTTTTCAAAATGTGTATAGCAGT 2148
Qy 753 yAlaProValThrValTrpMetAlaTyrAspThrGlyTyrThrGluAkrGlyTrpMetAspVa 773
Db 2149 GGCTCCAGTCTCCAGCTGGGAATATTACGCTGCTGTCTACACAGAGATTCATGGGTCT 2208
Qy 773 lPro-----GluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGl 791
Db 2209 CCCAACAAAGCATATATCTTGAGCACTATAAGAAATCAACTGTGTATGGCAAGACAGAGA 2268
Qy 791 uLysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAsnVa 811
Db 2269 ATATTTTCAGAAATGTAGAC-----TATCTTCTCATCCACGACAGATGATAATGT 2322
Qy 811 lHisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGl 831
Db 2323 GCATTTTCAAAACTCAGACAGATTTGTAAGCTCTGTTAATGCACAAGTGGATTTCCTCA 2382
Qy 831 nLeuGlnIleTyrProAsnGluArgHisSerIleArgCysProGluSerGly----- 848
Db 2383 GGCAATGTGGTACTCTGACACAGAACCCAGCGGCTTA-----TCCGGCCTGTCCAC 2430
Qy 849 -GluHisTyrGluValThrLeuLeuHisPheLeuGlnGlu 861
Db 2431 GAACCACTTATACACCCACATGACCCACTTCCTCTAAAGCAG 2470
RESULT 6
US-08-940-391-1
; Sequence 1, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D..
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2815 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-940-391-1
```

Alignment Scores:

Pred. No.: 5,43e-37 Length: 2815
Score: 426.00 Matches: 173
Percent Similarity: 39.08% Conservative: 106
Best Local Similarity: 24.23% Mismatches: 254
Query Match: 9.17% Indels: 182
DB: 2 Gaps: 31

US-09-976-674-3 (1-863) x US-08-940-391-1 (1-2815)

QY 195 ProAlaSerProAlaPhePheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIle 214
DB 734 CCAGGAGATCCACCTTTT
QY 215 GluThrGluGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAsp 234
DB 752 -----CAATAACAATTT-----AATGAAGAGAAATAAATAATTT 787
QY 235 AspProLysSerAlaGlyValAlaThrPheValIleGlnGluPheAspArgPheThr 254
DB 788 AAT-----GCAATCCAGACTGGGTTTATCAAGAGGAATG---CTTCCTACA 832
QY 255 GlyTyr-----TrpTrpCysProThrAlaSerTrpGluGlySerGluGlyLeuLysThr 272
DB 833 AATATATGCTCTCTGGTGGTCTCTAATGGAAATTT----- 868
QY 273 LeuArgIleLeuTyrGluGluValAspGluSerGluValGluValIleHisValProSer 292
DB 869 -----TTGGCATATCGGAAATTAATGATAAGGATATACAGTTATTCCTATTCCTAT 922
QY 293 ProAlaLeuGluGlu-----ArgLysThrAspSerTyrArgTyrProArgThrGlySerLys 311
DB 923 TATGGCGATGAACAATATCTAGAACAAATAAATATTCATACCAAGAGCTGGAGCTAAG 982
QY 312 AsnProLysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleVal 331
DB 983 AATCCCGTTGTCGGATA-----TTTATTATC 1009
QY 332 SerThrGlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysVal----- 349
DB 1010 GATACCACTTACCTCGGTATGAGTCCCGAGAGTGGCTGTTCCAGCAATGATAGCC 1069
QY 350 -----GluTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyrAlaTrpAla 366
DB 1070 TCAAGTGATTATTATTCAGTTGGCTCAGCTGGGTACTGATGAACGAGTATGT----- 1123
QY 367 MetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeuProProAlaLeuPhe 386
DB 1124 -----TTGCAGTGGCTAAAGAGCTCCAGAAATGTTTCGGTCTCTCT 1165
QY 387 Ile-----ProSerThrGluAsn----- 392
DB 1166 ATATGTCACTTCAGGGAAGACTGGCAGACATGGATTGTTCCAAAGACCCAGGAGCATATA 1225
QY 393 GluGluGlnArgLeu-----AlaSerAlaArgAlaValProArgAsnValGlnProTyr 410
DB 1226 GAAGAAAGCAGAACTGGATGGCTGGGTGATCTCTTCTTCAAGACCACTTTTCAGCTAT 1285
QY 411 ValValTyrGluGluValThrAsnValTrpIleAsnValHisAspIlePheTyrProPhe 430
DB 1285 ----- 1285
QY 431 ProGlnSerGluGlyGluAspGluLeuCysPheLeuArgAlaAsnGluCysLysThrGly 450
DB 1286 -----GATGCCATTTCTGACTACAAATAATTTAGTGACAGGATGGC 1327
QY 451 PheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGlu 470
DB 1328 TACAAACATATTCACTATATC----- 1348
QY 471 ProPheSerProGlyGluAspGluPheLysCysProIleLysGluGluIleAlaLeuThr 490
DB 1349 -----AAAGACACTGTGGAAATATGCTATTTCAAATTACA 1381

QY 491 SerGlyGluTrpGluValLeuAlaArgHisGlySerLysIleTrpValAsnGluGluThr 510
DB 1382 AGTGGCAAGTGGGAGGCCATA-----AATATATTCAGACGTAAACACAGAT 1426
QY 511 LysLeuValTyrPheGlnGlyThr---LysAspThrProLeuGluHisHisLeuTyr--- 528
DB 1427 TCACTGTTTATTCTAGCAATGAATTTGAAGAATACCTCGAAGAAGAAACATCTACAGA 1486
QY 529 -----ValValSerTyrGluAlaAlaGlyGluIleValArgLeuThrThrProGlyPhe 546
DB 1487 ATTAGCATTTGAAGCTATCTCTCAAGCAAGAAGTGTGTT----- 1525
QY 547 SerHisSerCysSerMet-----SerGlnAsnPheAspMet 558
DB 1526 -----ACTTGCCATCTPAAGAAAGAGGTGCAATATTACACAGCAAGTTTCAGGCAC 1579
QY 559 PheValSerHisTyrSerSerValSerThrProCysValHisValTyrLysLeu--- 577
DB 1580 TAGCCCAAGTACTATGCACCTGTCTGTACGGCCCGCAGGCATCCCAATTTCCACCTTCAT 1639
QY 578 SerGlyProAspAspAsp-ProLeuHis-----LysGlnProArgPheTrpAlaSerMe 595
DB 1640 GATGGAGCGCACTGATCAAGAAATTAATAATCCTGGAAGAAACAAAGCAATTTGGAATGCT 1699
QY 595 tMetGluAlaAlaSerCysProAspTyrValProProGluIlePheHisPheHisTh 615
DB 1700 TTGAAAAATATCCAGCTGCCT-AAAGAG-----GAAATTAAGAAACTTGAAGT 1746
QY 615 rArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLys 635
DB 1747 AGATGAATTAATTTATGTTACAGATGATCTCTCTCTCTCAA---TTTGACAGATCAAA 1803
QY 635 sLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSe 655
DB 1804 GAAGTATCCCTTCTAATTCAGTGTATGTTGGTCCCTGCAGTCAGAGTGTAGGTCTGT 1863
QY 655 rPhe-----LysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAl 673
DB 1864 ATTTGCTGTTAATTTGATATCTTATCTT-----GCAAGTAAGGAGGATGCT 1911
QY 673 aValValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLe 693
DB 1912 CATTTGCTTGGTGGTGGTGGAGCAGCTTTCCAAAGGTGACAAACTCTCTATGAGT 1971
QY 693 uLysAsnGlnMetGlyClnValGluIleGluAspGlnValGluClyLeuGlnPheValAl 713
DB 1972 GTATCCAAAGCTGGGTGTTTATGAAGTTGAAGACCAAGATTACAGCTGTCTGAAATTCAT 2031
QY 713 aGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTrpSerTyrGlyG 733
DB 2032 AGAA---ATGGGTTTCATTGATGAAAAAGAAATAGCCATATGGGCTGGTCTATGGAGG 2088
QY 733 yPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaG 753
DB 2089 ATACGTTTCATCATCTGGCCCTTCATCTCGAAGTGGTCTTTTCAAAATGTGCTATAGCAGT 2148
QY 753 yAlaProValThrValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspVa 773
DB 2149 GGTCTCAGTCTCCAGCTGGGAATATTACGGCTGTCTACACAGAGAGATTTCATGGGTCT 2208
QY 773 lPro-----GluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValG 791
DB 2209 CCCAACAAAGGATGATAATCTTCGACACTATAAGAAATTCACCTGTGATGCCAAGACAGA 2268
QY 791 uLysLeuProAsnGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVa 811
DB 2269 ATATTTTCAGAAATGTAGAC-----TATCTCTCATCCACGGAACACAGCATATAATGT 2322
QY 811 lHisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaClyLysProTyrG 831
DB 2323 GCACCTTTCAAAACTCAGCAGACAGATTGCTAAAGCTCTGGTTAATGCACAGTGGATTCCA 2382

Qy 831 nLeuGlnIleTyrProAsnGluArgHisSerIleArgCysProGluSerGly----- 848
 Db 2383 GGCAATGGGTACTCTGACCAAGACCGGCTTA-----TCCGGCCTGTCCAC 2430
 Qy 849 -GluHisTyrGluValThrLeuLeuHisPheLeuGlnGlu 861
 Db 2431 GAACCACTATACACCCACATGACCCACTTCTCTAAAGCAG 2470
 RESULT 7
 US-08-699-103B-1/c
 ; Sequence 1, Application US/08699103B
 ; Patent No. 6107462
 ; GENERAL INFORMATION:
 ; APPLICANT: Rine, Jasper D.
 ; APPLICANT: Hampton, Randolph
 ; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
 ; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 2200 Sand Hill Road, Suite 100
 ; CITY: Menlo Park
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/699,103B
 ; FILING DATE: 16-AUG-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/002,581
 ; FILING DATE: 17-AUG-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Green, Grant D.
 ; REGISTRATION NUMBER: 31,259
 ; REFERENCE/DOCKET NUMBER: 09272/005001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/322-5070
 ; TELEFAX: 650/854-0875
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4982 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 US-08-699-103B-1
 Alignment Scores:
 Pred. No.: 1.5e-25 Length: 4982
 Score: 327.50 Matches: 78
 Percent Similarity: 53.99% Conservative: 37
 Best Local Similarity: 36.62% Mismatches: 83
 Query Match: 7.05% Indels: 15
 DB: 3 Gaps: 7
 US-09-976-674-3 (1-863) x US-08-699-103B-1 (1-4982)
 Qy 637 HisProThrValLeuPheValTyrGlyProGlnValGlnLeuValAsnSerPhe 656
 Db 4941 TATCTGTATTTCCTTTCGATATGGGGACCGAATTCACCAAGTTGTCAAAACGGTT 4882
 Qy 657 Lys---GlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValVal 675
 Db 4881 TCCGTAGGATTTAATGAAGTG-----GTAGCTTCACAAATTAACCGCAATGTAGTT 4831
 Qy 676 ValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsn 695
 Db 4830 GTTGTGTACGGTCGTGGTACTGGCTTCAAGGTCAAGACTTTAGATCCCTTGTTCGGCAT 4771

Qy 696 GlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLys 715
 Db 4770 AGGCTCGGTATTACGAGCGCCGACCAAAATATCT-----GCGGCTTCCTTA 4723
 Qy 716 TyrGly-----PheIleAspLeuSerArgValAlaIleHisGlyTyrSerTyrGly 732
 Db 4722 TATGGTTCTTTAACTTTTGTGTGATCCGCAAAAGATTCTCTATTGTTGGTTCATACGGG 4663
 Qy 733 GlyPheLeuSerLeuMetGlyLeuIleHisLys---ProGlnValPheLysValAlaIle 751
 Db 4662 GGGTACCTGCACACTAAAAAATTTGGAGAAAGATGCGGAAGACATTTCAAAATACGGGATG 4603
 Qy 752 AlaGlyAlaProValThrValTyrMetAlaTyrAspThrGlyTyrThrGluArgTyrMet 771
 Db 4602 TCAGTTGCCAGCAGTAACCGACTGGAGATTTACGATTTCTGTTTATACAGAGGTACATG 4543
 Qy 772 AspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGlu 791
 Db 4542 CATACTCTCAAGAAACTTTTGATGATACGTAGTAATCAAGCGTT-----CATATATGTC 4489
 Qy 792 LysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnVal 811
 Db 4488 ACTGCTTTGGCACAAGCAAAATAGATTTTGTGTGATGCACGGAACAGGAGATGATAACGTT 4429
 Qy 812 HisPhePheHisThr---AsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyr 830
 Db 4428 CACTTCAAAATTCCTAAAGTTTCTGGACCTTTGGATCTAAATGGTGTGGAATAATAT 4369
 Qy 831 GlnLeuGlnIleTyrProAsnGluArgHisSerIleArg 843
 Db 4368 GACGTCCACGCTTTTCTCTGACTCAGATCATAGTATAAGA 4330
 RESULT 8
 US-09-229-059-1/c
 ; Sequence 1, Application US/09229059
 ; Patent No. 633172
 ; GENERAL INFORMATION:
 ; APPLICANT: Rine, Jasper D.
 ; APPLICANT: Hampton, Randolph
 ; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
 ; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 2200 Sand Hill Road, Suite 100
 ; CITY: Menlo Park
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/229,059
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/699,103
 ; FILING DATE: 16-AUG-1996
 ; APPLICATION NUMBER: 60/002,581
 ; FILING DATE: 17-AUG-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Green, Grant D.
 ; REGISTRATION NUMBER: 31,259
 ; REFERENCE/DOCKET NUMBER: 09272/005001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/322-5070
 ; TELEFAX: 650/854-0875
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4982 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-229-059-1

Alignment Scores:

Pred. No.: 1.5e-25 Length: 4982
Score: 327.50 Matches: 78
Percent Similarity: 53.99% Conservative: 37
Best Local Similarity: 36.62% Mismatches: 83
Query Match: 7.05% Indels: 15
DB: 4 Gaps: 7

US-09-976-674-3 (1-863) x US-09-229-059-1 (1-4982)

QY 637 HisProThrValLeuPheValTyrGlyProGlnValGlnLeuValAsnSerPhe 656
Db 4941 TATCCTCTATTTTCTTTTGCATATGGGGACCGGAATTCCTCAACAAGTTGTCAAAACGTTT 4882
QY 657 Lys---GlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValVal 675
Db 4881 TCCGTAGATTATTAAGTG-----GTAGCTTCACAAATTAACGCAATTTGTAGTT 4831
QY 676 ValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsn 695
Db 4830 GTTGTTCACGCTGTGTGCTTCAAAAGTCAAGACTTTAGATCCCTTGTTCGCGAT 4771
QY 696 GlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLys 715
Db 4770 AGGCTCGGTGATTACGAGGCCCGCGCAAAATATCT-----GCGGCTTCCTTA 4723
QY 716 TyrGly-----PheIleAspLeuSerArgValAlaIleHisGlyTyrPserTyrGly 732
Db 4722 TATGTTCTTTAACTTTTGTGATCCGCAAAAGATTTCTTATTGTGTGTCATACGG 4663
QY 733 GlyPheLeuSerLeuMetGlyLeuIleHisLys---ProGlnValPheLysValAlaIle 751
Db 4662 GGGTACCTGCACCTAAAACTTTGGAGAAAGATGGCGGAAGACATTTCAAAATACGGGATG 4603
QY 752 AlaGlyAlaProValThrValTrpMetAlaTyrAspThrGlyTyrThrGluAqrTyrMet 771
Db 4502 TCAGTTGGCCAGTAACCGACTGGAGATTTTACGATTTCTGTTATATCTAGAGGTACATG 4543
QY 772 AspValProGlnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGlu 791
Db 4542 CATACTCTCAAGAAACTTTGTGATAGTACGTAGTAATCAACGGTT-----CATAAATGTC 4489
QY 792 LysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnVal 811
Db 4488 ACTGCTTTGGCACAAGCAATAGATTTTGTGATGCACGGAACAGGAGATGATAACGTT 4429
QY 812 HisPhePheHisThr---AsnPheValSerGlnLeuIleArgAlaGlyLysProTyr 830
Db 4428 CACTTCAAAATCCCTAAAGTTTCTGGACCTTTTGGATCTAAATGCTGTGGAATAATTAT 4369
QY 831 GlnLeuGlnIleTyrProAsnGluArgHisSerIleArg 843
Db 4368 GAGCTCACGCTCTTCTGCTAGCTACATAGTATAAGA 4330

RESULT 9

US-09-221-017B-253

; Sequence 253, Application US/09221017B
; Patent No. 6444799

GENERAL INFORMATION:

; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESS: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA

; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PPI182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PPI546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PPI2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: MONROY, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141

INFORMATION FOR SEQ ID NO: 253:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...543
US-09-221-017B-253

Alignment Scores:

Pred. No.: 3.73e-26 Length: 543
Score: 318.00 Matches: 60
Percent Similarity: 57.89% Conservative: 39
Best Local Similarity: 35.09% Mismatches: 68
Query Match: 6.84% Indels: 4
DB: 2 Gaps: 2

US-09-976-674-3 (1-863) x US-09-221-017B-253 (1-543)

QY 667 LeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySerCysGlnArgGly 686
Db 43 ATGGCACAGAAAGCGCTATGCGTCTTTACGGTGGATAGTCGCGGATCGCAATAGAGG 102
QY 687 LeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGluAspGlnVal 706
Db 103 GCTGCTTTTCGAGCAGGTTATTTCATCGTCTGTTGGGCGAGACCGAGATGCGCGATCAGATG 162
QY 707 GluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgValAlaIle 726
Db 163 TCGGCTGTGGATTTCCTC---AAGAGCCAAATCATGTTGGATGCCGATAGAGGATG 219
QY 727 HisGlyTyrPserTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnVal 746
Db 220 CATGGCTGGAGCTATGTTGGCTTTATGACTACGAACTGATGCTTACGCGCGGATGTC 279

Qy 747 PheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyrAspThrGlyTyr 766
Db 280 TTTAAAGTCGGAGTCAGCGGGGCTTCATAGACTGGAATCGATATGAGATTATGAC 339
Qy 767 ThrGluArgTyrMetAspValProGluAsnGlnHisGlyTyrGluAlaGlySerVal 786
Db 340 GGTAGCGCTTATTCGATGCCACGAAATCCCGAAGGATGATGCTGCCAACCTG 399
Qy 787 AlaLeuHisValGluLysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPhe 806
Db 400 -----CTCAACGAGCGGTGATCTGAAAGGACGACATATGCTGATTCATGGACGG 450
Qy 807 LeuAspGluAsnValHisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAla 826
Db 451 ATCGATCCGGTCGTGGTATCGCAGCATTCACCTCTTTCTTGTATGCTTGGTGAAGCA 510
Qy 827 GlyLysProTyrGlnLeuGluIleTyrProAsn 837
Db 511 CGCACCTATCTGACTTACGTCATCCGAGC 543
RESULT 10
US-09-221-017B-646/c
; Sequence 646, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 646:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN

; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...657
US-09-221-017B-646
Alignment Scores:
Pred. No.: 6.07e-24 Length: 657
Score: 299.50 Matches: 71
Percent Similarity: 50.53% Conservative: 24
Best Local Similarity: 37.77% Mismatches: 84
Query Match: 6.45% Indels: 9
DB: Gaps: 4
US-09-976-674-3 (1-863) x US-09-221-017B-646 (1-657)
Qy 605 TyrValProGluIlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMet 624
Db 566 TACAACCCGGAAGGAGTTTACCACCTATCAAACTCAATCGGGCTTGAACCTGAATGCCTGG 507
Qy 625 IleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyr 644
Db 506 ATCGTGAAGCCTATTGATTTCGATCCCTCTCGCCACTATCTCTCTGATGGTACAGTAT 447
Qy 645 GlyGlyProGlnValGlnLeuVal-----AsnAsnSerPheLysGlyIleLysTyrLeu 662
Db 446 AGCGGTCCCACTCCACGAGGTATGGATCGCTATTCTATTCGATGGGAACACTAC--- 390
Qy 663 ArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySer 682
Db 389 -----CTTGTCATCGAAAGGTTACGTCGTGCATGTGTGGATGGGCGGTGGCACC 342
Qy 683 CysGlnArgGlyLeuArgPheGluAlaLeuLysAsnGlnMetGlyGlnValGluIle 702
Db 341 GGTGCTCGCGCGAAGATGGCGCAAGTGCTACCTACATCGCACTCGGTGTATTCGAAAGC 282
Qy 703 GluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSer 722
Db 281 GATGATCAGATAGCAGCGCCACTCTATAGGACAG---CTGCCCTATGTGGATCAGCT 225
Qy 723 ArgValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHis 742
Db 224 CGTATCGCATATGGGGTGGAGCTATGGCGCTATACCACTAATGAGTTTGTGTGG 165
Qy 743 LysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyr 762
Db 164 GGAATGTTATCATCAAGCGGGGATGCGCTTCTCTGGCAGACTGGCGGTTTCTTAC 105
Qy 763 AspThrGlyTyrThrGluArgTyrMetAspValProGluAsnGlnHisGlyTyrGlu 782
Db 104 GATTGCGTTTACACCGAAGCTTTCATGCTGCTGGCAGACTGGCGGTTTCTTAC 105
Qy 783 AlaGlySerValAlaLeuHisVal 790
Db 44 ATGCTCTCT---GCTCTTGATGTG 24
RESULT 11
US-09-392-184-31/c
; Sequence 31, Application US/09392184
; Patent No. 6395889
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: 5800-55
; CURRENT APPLICATION NUMBER: US/09/392.184
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 612
; TYPE: DNA

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(612)
; OTHER INFORMATION: prolyl oligopeptidase)
; NAME/KEY: misc_feature
; LOCATION: (1)..(612)
; OTHER INFORMATION: n = A,T,C or G
US-09-392-184-31

Alignment Scores:
Pred. No.: 2,31e-20 Length: 612
Score: 267.00 Matches: 71
Percent Similarity: 57.59% Conservativeness: 20
Best Local Similarity: 44.94% Mismatches: 65
Query Match: 5.75% Indels: 4
DB: 4 Gaps: 0

US-09-976-674-3 (1-863) x US-09-392-184-31 (1-612)
Qy 701 GluIleGluAspGlnValGluGlyLeu-GlnPheValAlaGluLysTyrGlyPheIleAs 720
Db 519 GAATGTGCGATCAGTGGGNGGACTCCCAATATCTACTCTCGAATGATTTTCATGA 460
Qy 720 pLeu-SerArgValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyL 740
Db 459 CNTAAGATCGTGTGGCAATCCNCGCTG-TCCATGAGGATACCTCTCCCTGATGGCAN 401
Qy 740 eulleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpM 760
Db 400 TAATTTCAGAGTCAGATATCNCAGGTTCT-ATTTCGGGGCCCACTCACTCTGTGGA 342
Qy 760 etAlaTyrAspTyrGlyTyrThrGluArgTyrMetAspValProGluAsnGlnHisG 780
Db 341 TCTTCTATGAACAGGAACNCGGAGGATTANATGGTCACCTNNCCAGNATGGACAGG 282
Qy 780 lyTyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgL 800
Db 281 GCTATTACTTAGGATCTGTGGCCATTCCANCAGGAAAGTTCCCTCTGNACCCCAATNGT 222
Qy 800 euLeuIleLeuHisGlyPheLeuAspGluAsnValHisPheHisThrAsnPheLeuV 820
Db 221 TNCCTTCTTANATGGTTCTCTGGATTAGGATTTCCANTTTTCANACCAGNANANTAC 162
Qy 820 alSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgH 840
Db 161 TNAGGTTTTAGTGAGGCTGGGAAACCAATGANTTNCAGAACCACTCNGGAGAAGN 102
Qy 840 isSerIleArgCysProGluSerGlyGluHisTyrGluValThrLeuLeu 856
Db 101 CCAACANNAGGNTCTGAATCGGGAGGACAATATNTGACCTNNACCTT 52

RESULT 12
US-09-016-080-2
; Sequence 2, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1896

; TYPE: DNA
; ORGANISM: Pyrococcus horikoshii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1896)
US-09-016-080-2

Alignment Scores:
Pred. No.: 1,5e-16 Length: 1896
Score: 240.50 Matches: 161
Percent Similarity: 35.97% Conservativeness: 107
Best Local Similarity: 21.61% Mismatches: 283
Query Match: 5.18% Indels: 195
DB: 3 Gaps: 35

US-09-976-674-3 (1-863) x US-09-016-080-2 (1-1896)
Qy 132 LysArgLeuGlyValPheGlyIleThrSerTyrAspPheHisSerGlu-----Ser 148
Db 67 AAGGAAAGCGGTCTTCAAGTAACGAGATAAGCCTCAAGACGATGATTACTTCTCT 126
Qy 149 GlyLeuPheLeuPheGlnAlaSerAsnSerLeuPheHisCysArgAspGlyGlyLysAsn 168
Db 127 AAGCTTTTACCTCTAC-----GATGGAAGAGG--- 153
Qy 169 GlyPheMetValSerProMetLysProLeuGluIleLysThrGlnCysSerGlyProArg 188
Db 154 -----GTAAACCTTCACTCAGGGAACAGGATTTCTAATCCAAGG 195
Qy 189 MetAspProLysIleCysProAlaAspProAlaPhePheSerPheIleAsn----- 205
Db 196 TTCTCTCCA-----AATGGGAAGCTTATAGCATTTACCTCAAGAGGAT 240
Qy 206 -----AsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg 221
Db 241 AAGGAAGAAAGGAATCAGAGCTCTACGTGATTCCACGATGGGAGAGGCCAGACTT 300
Qy 222 LeuThrPheCysHisGlnGlyLeuSerAsnVal-----LeuAspAspProLysSerAla 239
Db 301 TTAGCAAGTTCANATACGGGATAAAGAACCTGCGCTTACCAGAGATGGGAAGATATA 360
Qy 240 GlyValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpCys 259
Db 361 GCGGTGCTTACCCCTATAGACGTTGAGAAA----- 390
Qy 260 ProThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGlu 279
Db 391 -----AAAGGAATGATGACGTTCCACATTATAAG----- 420
Qy 280 ValAspGluSerGluValGluValIleHisValPro-SerProAlaLeuGluGluArgLy 299
Db 421 -----GAATACCATCTCGTTTAAATGGAG----- 445
Qy 299 sThrAspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLe 319
Db 445 ----- 445
Qy 319 uAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuVa 339
Db 446 -----TTGGCTGGATCTACGGAAGAAAGAAAGCGTT 474
Qy 339 lGlnProPheSerSerLeuPhePheProLysValGluTyrIleAlaArgAlaGlyTrpThr 359
Db 475 GTCTACCTCTTGTGACGTTG-----AGACGGGAAGAAAG 510
Qy 359 gAspGlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuVa 379
Db 511 AGAC-----TAACTCCAAAGAAACCTAAATGTTGAT-CAGATAAG 548
Qy 379 lLeuLeuProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArg---LeuAl 398
Db 549 GTTCCACACAGGTTAGACTATATCTACGGCCCAAGAGGATAGGAAGAAAGAAACCTCTGAT 608
```

QY 398 aserAlaArgAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAs 418
 Db 609 ATCCGATCTTTACGCTCTCGAGATAGAAA-----GTTAGGAAGCTGACCCAGG 659
 QY 418 nvalTrpIleAsnValHisAspIlePheTyrProPheProGlnSerGluGlyGluAspG1 438
 Db 660 GAAGTGG---AGGATACATCGAC---TTCCTCCCTT-----GATGACGG 698
 QY 438 uLeuCysPheLeuArgAlaAsnGluCysLysThrGlyPhe-----CysHisLeuTy 455
 Db 699 AAGCTTCGTACTTAAAGGCTACACTTAGAAGGGGATCCCAACCAAGCCCACTCTA 758
 QY 455 rLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProG1 475
 Db 759 CCAC-----TACGATCCAA 773
 QY 475 yGluAspGluPheLysCysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpG1 495
 Db 774 GACAGGAGAACTTAAGAGGCTCAACAAGGAT-----TTAGACAGGAACGCTTACAA 824
 QY 495 uValLeuAlaArgHisGlySerLysIleTrpValAsnGluGluThrLysLeuValTyrPh 515
 Db 825 CTCCTTA-----AACTCCGATGTCGAGGAAGTCAGAGGCCGAGCTTGTGTACAA 875
 QY 515 eGlnGly-----ThrLysAspThrProLeuGluHisLysLeuTyrValVa 530
 Db 876 GGAGGGGTGGATCTACTATGCGCAACGATGGCCCT---AGGCCAAACCTCTTTAGGGT 932
 QY 530 lSerTyrGluAlaIleValArgLeuThrProGlyPheSerHisSerCy 550
 Db 933 CAACCTTAGAT-----GGAAGATTGAAGGGTAATA-----GGTGAGATAGAAGCGT 980
 QY 550 sSerMetSerGlnAsnPheAspMetPheValSerHisTyrSerValSerThrProPr 570
 Db 981 TGAAGCTTCGATATAGGGGATACATAGCTTCCAGCGCTCAAGATCTGTAAACCCCAAC 1040
 QY 570 oCysValHisValTyrLysLeuSerGlyProAspAspProLeuHisLysGlnProAr 590
 Db 1041 TGAGCTCTACATATACAGG-----GATGGAAGGAGGAAGGTTACCGA 1085
 QY 590 gPhe-----TrpAlaSerMetMetGluAlaAlaSerCysProProAspTyrValProPr 608
 Db 1086 CTTTAAACAAATGGATAAAGGTTACACCCCTTTCAAAA-----CC 1124
 QY 608 oGluIlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysPr 628
 Db 1125 TGACACTTTAAGCTTAAAGCAAGTACGGGGTTGAAATAGATCCCTGGGTAATGAACCC 1184
 QY 628 oHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyrGlyGlyProG1 648
 Db 1185 GGTGAACCTCAGGAAGAAAGAAAGTATCCAGCTATTCTAGAGATCCACGGTGGCTCTAA 1244
 QY 648 nValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeu---ArgLeuAsnThrLe 667
 Db 1245 AACCGCTTAC-----GGTTACGCTTTTATGACGAGTTCACCGAGTTCCACGTTT 1286
 QY 667 uAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySerCysGlnArgGlyLe 687
 Db 1287 AACCTCTAAGAGCTTCGTGATATCTCAAAATCTTACAGGGAGCGATGGCTACGGAGA 1346
 QY 687 uArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGluAspGlnValG1 707
 Db 1347 GGAGTTC---CCGATATAGGGGACACACTATGGGAGAGGAGGATTACAGGATTTAATGGA 1403
 QY 707 uGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgValAlaIleH1 727
 Db 1404 GGTAGTCGATGAACCATTAAGGAGATTGACTTTCATAGTGGGAAAGGCTAGGAGTTAC 1463
 QY 727 sGlyTrpSerTyrGlyPheLeuSer-----LeuMetClyLeuIleHisLysProG1 745
 Db 1464 CGGGGGTTCCTTATGGCTTTCAGCAACTGGATATAGCTCGGACATACCAACAGG----- 1518
 QY 745 nValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyr----- 762

Db 1519 ---TTCAAGCCGCTGTAAACCCAGAGATCAATTTCAAATTTGGATAAGCTTCTTCGGGAC 1574
 QY 763 ---AspThrGlyTyr-----ThrGluArgTyrMetAspValProGluAsnAsnG1 778
 Db 1575 AACGGATATAGGTATTACTTTGCTCCAGATCAATAGAAAAGATCCCTCGAGCAACTT 1634
 QY 778 nHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAs 798
 Db 1635 GGAAGTTATTGGGAAAAGAGC-----CCATTAAAGTACGCTCCCAACGTTGAAAC 1685
 QY 798 nArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPh 818
 Db 1686 TCCCTGCTTATATCCACTCTACCGAGACTACAGGTGTTGGCTTCCCGAGGCATTGCA 1745
 QY 818 eLeuValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnG1 838
 Db 1746 ACTCTTCATATCCCTAAATACCTGGGGAAGAGAGTTGAATTGGCAATATTATCCAGGAGA 1805
 QY 838 uArgHisSerIle 842
 Db 1806 AATCATGACCTA 1818
 RESULT 13
 US-09-221-017B-1010/c
 : Sequence 1010, Application US/09221017B
 : Patent No. 6444799
 : GENERAL INFORMATION:
 : APPLICANT: ROSS, BRUCE C.
 : TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
 : NUMBER OF SEQUENCES: 1120
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: MORRISON & FOERSTER
 : STREET: 755 PAGE MILL ROAD
 : CITY: Palo Alto
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94304-1018
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: Windows
 : SOFTWARE: FASTSO for Windows Version 2.0b
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/221.017B
 : FILING DATE: 23-DEC-1998
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PP1182
 : FILING DATE: 31-DEC-1997
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PP1546
 : FILING DATE: 30-JAN-1998
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PP2911
 : FILING DATE: 09-APR-1998
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PCT/AU98/01023
 : FILING DATE: 10-DEC-1998
 : ATTORNEY/AGENT INFORMATION:
 : NAME: MONROY, Gladys H
 : REGISTRATION NUMBER: 32,430
 : REFERENCE/DOCKET NUMBER: 27340-20021.00
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 650-813-5600
 : TELEFAX: 650-494-0792
 : TELEX: 706141
 : INFORMATION FOR SEQ ID NO: 1010:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 3085 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: circular

```

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...3085
; US-09-221-017B-1010

Alignment Scores:
Pred. No.: 1 68e-13 Length: 3085
Score: 216.50 Matches: 139
Percent Similarity: 36.25% Conservative: 97
Best Local Similarity: 21.35% Mismatches: 262
Query Match: 4.66% Indels: 155
DB: Gaps: 21

US-09-976-674-3 (1-863) x US-09-221-017B-1010 (1-3085)
QY 318 LysLeuAlaGluPheGlnThrAspSerGlnGlyLysValSerThrGlnGluLysGlu 337
    ||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
Db 2888 AAGGCTTCGCTTTTCAGCTAAAGCCCAAGCGGCAAGTACCTCTCATACATGGAAGGAC 2829
QY 338 -----LeuValGlnProPheSer-----Ser 344
    ||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
Db 2828 GATGCGGCAAGCCGATGCTATGTCAGGAAATTTGCCACCGGACCGTCAAGCGGTGCC 2769
QY 345 LeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrArgaspGlyLysTyrAla 364
    : : : ||| ||| ||| ||||| ||| : : : ||| : : : ||| : : :
Db 2768 ATCGAAGAAAGGACCAAGTATGTCAGGCTACGAGTGGATGATCAACGACGACGCTCTTC 2709
QY 365 TrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProAla 384
    : : : ||| ||| ||| ||||| ||| : : : ||| : : : ||| : : :
Db 2708 TTGTGTCATG----- 2700
QY 385 LeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAlaArgAlaValPro 404
    ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
Db 2699 -----GACAAAGAGGAGGAATGAGAACTATACCTCTTTTTCGAATATCGAGGC 2649
QY 405 ArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTrpIleAsnVal--- 423
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2648 AGCAATACCGCGCATCTACCCCTTTGAGGAGTGAAGCTTTCGATCCTCAACATGCTC 2589
QY 424 -----HisAspIlePheTyr----- 428
    ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
Db 2588 AAAGACGAGGACATACATGATCATATCCATGACAAACAAACATCCGCGATCTTCGAA 2529
QY 429 ProPheProGlnSerGluGlyGluAspGluLeuCysPheLeuArgAlaAsnGluCysLys 448
    ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
Db 2528 CCTACAAACTGAATGTAGTAAACAGCGGAGCTGACCCAGCTCTACGAGATAAG----- 2475
QY 449 ThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTrp 468
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2474 -----GATGCGGCAACCCCATTCAGGTTACGAGTTC 2442
QY 469 SerGluProPheSer---ProGlyGluAspGluPheLysCysProIleLysGluIle 487
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2441 GACAAGGAGCGGCAACTGCGTGATACAGCGGCTCTGTAACCGGATCGAATCCGAGTTG 2382
QY 488 -----AlaLeuThrSerGlyGluTrpGluValLeu-AlaArgHisGlySerLysII 504
    ||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
Db 2381 TACTACAAGGATTTGGCTACGGGCGAGTTCCGCTCTGCTGAAGAAACAC---ACTGGGAC 2325
QY 504 eTrpValAsnGluThrLysLeuValTyrPheGlnGlyThrLysAspThrProLeuG1 524
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2324 GACACCTTCGGATCATCGGTTCACTATGCTCCAAACAAACAGACGAGCCTATGTA 2265
QY 524 uHisHisLeuTyr-----ValValSerTyrGlu----- 533
    : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
Db 2264 CTGACCAACCTGGACGACGACGACGACTCGTATCGTACTCTACGACCTGAAGCAGAACAG 2205
QY 534 -----AlaAlaGlyLulleVa 539
    : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :

```

```

Db 2204 ATCATCCGCGAGATCTTCGCCAACGAAGACTAGACGCTACGCGCCTGCACCTCTCTCGT 2145
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 539 lArgLeuThrThrProGlyPheSer----- 547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2144 AAGAGAAACTACGAATAGACCTCATGGCCCTACGAAGCGGAGAGTCCGCTAGTGTACCC 2085
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 -----HisSerCys-----SerMetSerGlnAs 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2084 GTAAGTGCCACCTACAAAGAGCTGCACAAGCTGATGAAAAGAAATTCAAAGGGCAAGAA 2025
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 555 nPheAspMetPheValSerHisTyrSerSerValSerThrProCysValHisValty 575
    | : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 2024 TTCTCCGTGGTCGATT-----ACGATGATGATGAGACCATCTCTGTATCGCGTA 1974
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 575 rLys-----LeuSerGlyProAspAspProLeuHisLysGlnProArgPheTr 592
    ||||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1973 CAAAGGCACAAAGCTATACGCACTTACTACCACTTCCAGTTCGATA---CGCCACCAAGAAGTTT 1917
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 592 pAlaSerMetMetGlu-----AlaAlaSerCysProProAspTyrValProGluI1 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1916 ACCCTCTCTATGACCTGATGCTCAGCTCAAG-GAGGAAGATATGCCCGAGATGCCGCC 1858
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 610 ePheHisPheHisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAl 630
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1857 CATCAAAATTCAGAGCGCGGACGAGCTCACTATCATCTGGCTTTTATCACTCTGCCGAAGC 1798
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 630 aLeuGlnProGlyLysLysHisProThrValLeuPheValTyrGlyProGlnValG1 650
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1797 AGCCCTCGAAGGAAGAAAGTACCCCTGCTCAATCCGATGAGGAGGCCCCCAA----- 1743
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 650 nLeuValAsnAsnSerPheLysGlyIleLys-----TyrLeuArgLeuAs 665
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1742 -----GGCATACCGGACTCATGGGCTTCAATCCCGAGACCCA 1705
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 665 nThrLeuAlaSerLeuGlyTyrAlaValValValLeuAspGlyArgGlySerCysGlnAr 685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1704 GCCTCTCGCGAGCGGATATGCCACCTGCAAGTCAATTTCCGCATCTCAGCGGGATA 1645
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 685 gGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluLysG1 705
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1644 CGCAAGGAATCTCTCGTCCGGATTCACACAGATCGGTGCGCAAGCCATGACGACGTG 1585
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 705 nValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgValAl 725
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1584 GGAGGACGGTGTGCGCTAT---GCTATACGCCAAGGTGGGTGGATCTCAGACGATCGC 1528
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 725 alIleHisGlyTrpSerTyrGlyPheLeuSerLeuMetGlyLeuIleHisLysProG1 745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1527 CATATACGGTCCGACGACGCTGTTATGCCAGCTGATGGGTCTGGTGAACACCCGA 1468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 745 nValPheLysValAlaIleAlaGlyAlaProValThr----- 757
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1467 TCCTATGCTCGGCTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1408
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 758 -----ValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspValPr 774
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1407 CTTCGCCAGAAATTTGGAAGCGGTTTAAAGAAATGGTCAAGGAAATTTGGTACGACCTCGA 1348
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 774 oGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeuPr 794
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1347 -CAATCCGAGGAAGACGATATCGCAAGGAAGTGTCCCTCTTCCAGATCGACAAGA 1289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 794 o-AsnGluProAsnArgLeuLeuIleHisGlyPheLeuAspGluAsnValHisPhep 814
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1288 TCATTAAGCCA-----CTGTTCTGCTACAGGAGGACCAAGCCCGCGTGAATATCA 1235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 814 heHisThrAsnPheValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnI 834
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1234 ACGAGTCCGATCATAGTCAACGCGCATGCGTCCCGCGGATTCGAAGTACCCATATGG 1175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 834 leTyrProAsnGluArgHisSerIleArgCysProGluSerGlyGluHisTyrGluValT 854
    : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
Db 1174 TGAAGTACAACGAGCGCACCGATTCCATCGTGAAGAAAACTCCATGGAGCTATACCGTG 1115
Oy 854 hrLeuLeuHisPheLeuGlnGluTyrLeu 863
Db 1114 CCAATGCTCGGTTTCTTCGCCAAACACCTG 1086
RESULT 14
US-09-280-116-172
; Sequence 172, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280.116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 172
; LENGTH: 502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: prolyl oligopeptidases
US-09-280-116-172
Alignment Scores:
Pred No.: 1.61e-12 Length: 502
Score: 195.50 Matches: 47
Percent Similarity: 55.38% Conservative: 25
Best Local Similarity: 36.15% Mismatches: 54
Query Match: 4.21% Indels: 4
DB: 4 Gaps: 2
US-09-976-674-3 (1-863) x US-09-280-116-172 (1-502)
Oy 732 GlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIle 751
Db 1 GGTGGCTATATGCATCAATCATCTTAAATCAGATGAAGCTTTTAAATGTGGATCC 60
Oy 752 AlaGlyAlaProValThrValTyrMetAlaTyrAspThrGlyTyrThrGluArgTyrMet 771
Db 61 GTGGTTGCACCTATCACAGACTTGAATTTGTATGCTCTCTCTGAAAGATACCTT 120
Oy 772 AspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGlu 791
Db 121 GGGATGCCATCTAAGGAAGAAGACACTTACCAGGCGCAGCTGTGCTACATAATGTTCA 180
Oy 792 LysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnVal 811
Db 181 GGCITGAAGAAGAA-----AATATATTAATATTCATGGAAGCTGCACACAAAAGTT 234
Oy 812 HisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGln 831
Db 235 CATTTCACACACTCAGCAGAATTAATCAAGCACCTAATAAAGCTGGAGTGAATTAATCT 294
Oy 832 LeuGlnIleTyrProAsnGluArgHisSerIleArgCysProGlu-SerGlyGluHisTyr 851
Db 295 ATGCAGGTCTACCCAGATGAAGGTCAATAAC--CGTATCTGAGAAGAGCAAGTATCATCT 351
Oy 851 rGluValThrLeuLeuHisPheLeuGln 860
Db 352 CTACAGCACATCTCAAAATTTCTCAAG 379
RESULT 15
US-08-227-689-1
; Sequence 1, Application US/08227689
; Patent No. 5521081
; GENERAL INFORMATION:
; APPLICANT: Tetsuya INAKA et al.
; TITLE OF INVENTION: DNA Coding For Enzyme
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,689
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/917,344
; FILING DATE: July 23, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHEetical:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Flavobacterium meningosepticum
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE: plasmid pFPEP03
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION: 1 to 259
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /note= "promoter region"
; FEATURE:
; NAME/KEY:
; LOCATION: 260 to 316
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /note= "signal sequence"
; FEATURE:
; NAME/KEY:
; LOCATION: 317 to 2374
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /note= "mature"
; OTHER INFORMATION: prolyl oligopeptidase coding region"
; FEATURE:
; NAME/KEY:
; LOCATION: 2375 to 2377
; IDENTIFICATION METHOD:
```

```

; OTHER INFORMATION: /note= "stop codon"
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
; US-08-227-689-1

```

```

Alignment Scores:
Pred. No.: 1.39e-07 Length: 2636
Score: 162.50 Matches: 124
Percent Similarity: 39.83% Conservative: 107
Best Local Similarity: 21.38% Mismatches: 224
Query Match: 3.50% Indels: 125
DB: 1 Gaps: 34

```

US-09-976-674-3 (1-863) x US-08-227-689-1 (1-2636)

```

QY 267 SerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluValValAspGluSerGluValGlu 286
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 593 AATGATGCTCTCAGCGCAATCTGTACTTTACAGAAAGATGCGCAGGTAAAGACGGAA 652

QY 287 ValIleHisValProSerProAlaLeuGluGluArgLysThrAspSerTyrArgTyrPro 306
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 653 GTATTTTGTAGATCCTAAT---AAGTTTCGGAAGAAAGAACCACTTCTCTGCGAAGTGT 709

QY 307 ArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAlaGluPheGlnThrAspSer 326
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 710 TCITTTTAAAGAAAGAACTCTGGTCGTATAGTATATACAGAGGAGGTTCGGACTGG 769

QY 327 GlnGlyLysIleValSerThrGlnGlu-----LysGluLeuValGln----- 340
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 770 AATAAGATTATTCTCGATCGGAAACCAAAAGCAACTTGATGAAACTCTATTGGAT 829

QY 341 ---ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrPheArg 359
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 830 GTTAAGTTTCAGTGGAAAT-----TCATGGTTGGGA 859

QY 360 AspGlyLysTyrAlaTyrAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuVal 379
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 860 GATGAGGATCTTTTATCCAGCTAT---GATAAGCCAAAGAAAGAAAGCAACGCTACTTCC 916

QY 380 LeuLeuPro-----ProAlaLeuPhe-----IleProSerThrGluAsnGluGlu 394
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 917 GGGATGACAGATAAACACAAAGTTTATTTCATAAGTTAGGAACGAAGCAGTCTCAGGAT 976

QY 395 GlnArgLeuAlaSerAlaArgAlaValProArgAsn---ValGlnProTyrValValTyr 413
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 977 GAATTGATTATGCGGGGTGATAAATTTCCAAGAAAGATATATAGGAGCTTTATGTAACCGAT 1036

QY 414 GluGluValThrAsnValTyrIleAsnValIleAspIlePheTyrProPheProGlnSer 433
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1037 GATCAG-----AGATATCTGGTGGTTTCGGCTGCAATGCA 1072

QY 434 GluGlyGluAspGluLeuGluCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysHis 453
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1073 ACCAAGCAACGAGCTT---TACATTAAAGACCTGAAGAATAAAACAGATTTTATTCG 1129

QY 454 LeuTyrLysValThrAla-----ValLeuLysSerGlnGly----- 465
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1130 ATT-----ATTACAGGTTTGTAGCAATGTAATGTTGCAGATACCCGAGCTGATACG 1183

QY 466 ---TyrAspTrpSerGluProPheSerProGlyGluAspGluPheLysCysProIleLys 484
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1184 CTTTATTGTTCCCGNATAAGATGCACCCGAATAAGCGACTGGTAAACACACGATTTCAG 1243

```

```

QY 485 GluGluIleAlaLeuThrSerGlyGluTrp---GluValLeuAlaArgHisGlySerLys 503
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1244 AATCCAAAGCGGAACA-----TGCAAGATGTGATTGCTGAACACCGCAACCA 1294

QY 504 IleTrpValAsnGluGluThrLysLeuValTyrPheGlnGlyThr-----LysAspThr 521
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1295 TTCCAATCAAT-----ACGGGAGCGGTTATTCTTCTTCTATATGAAGATGCA 1348

QY 522 ProLeuGluHisLeuTyrValValSerTyrGluAlaAlaGlyGluIleValArg--- 540
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1349 ATCGATCAG-----GTAAGCAATATGATAAACGGAAGCTTGTAAAGGCT 1396

QY 541 LeuThrThrProGlyPheSerHisSerCysSerMet-----SerGlnAsnPheAsp 557
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1397 ATAAATATACCGGAGTGTGTAATGCAAGCGGTTTGGGGGTGAAAAACGGAAGGAT 1456

QY 558 MetPheValSerHisTyrSerSerValSerThrProCysValHisValTyrLysLeu 577
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1457 CTGTATTACTCT---TTCACCAATATATTACGCGCAACGATCTTTAAATATAATGTA 1513

QY 578 SerGlyProAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGlu 597
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1514 ACAACAGGTAACTCTGAA---GTTTACCAGAGCCGAAAGTGAAGTCAAT----- 1561

QY 598 AlaAlaSerCysProAspTyrValProGluIlePheHisPheHisThrArgSer 617
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1562 -----CCGGAATAATTATGTTTCGAGCAGGTA-----TTCATATCTCATCT 1603

QY 618 Asp---ValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLys 636
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1604 GACGGGACTAAGATTCGATGATCAGCTACAGAAAGCCCTGAAAAAGACGGTAAA 1663

QY 637 HisProThrValLeuPheValTyrGlyGly-----ProGlnValGln 650
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1664 AACCTACATATATATACAGCTACGAGGAGTTAATATCAGCTCTTCAGCCTGCTTCTCT 1723

QY 651 LeuValAsnAsnSerPheLys-----GlyIleLysTyrLeuArgLeuAsnThrLeu 667
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1724 GTTGTAAATCAATCTGGATGGAACGGTGGTATT----- 1759

QY 668 AlaSerLeuGlyTyrAlaValValValIleAspGlyArgGlySerCysGlnArgGlyLeu 687
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1760 -----TATGCTGTTCGCAATATC-----CGTGGTGTGGAGAAATATGGTAAG 1801

QY 688 ArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGlu 707
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1802 AATGGCATGATCCCGGAACATAATCCACAAAAAGATGATTTAATGACTTTATTGCA 1861

QY 708 GlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHis 727
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1862 GCCGAGAGTACTTA---CAGAAAAACGGTTATACATCTAAGGAATATATGGCCTTTC 1918

QY 728 GlyTrpSerTyrGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPhe 747
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1919 GGACGTCTCAACGCGGCTCTTCTGTAGGGGCTACGATGACAATGCGCCCTGATTTGGCA 1978

QY 748 LysValAlaIleAlaGlyAlaProVal----- 756
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1979 AAGTTGCAATCCCGGAGTAGGACTGCTGGATATGCTTGTATATAAGTTTACAGCT 2038

QY 757 ---ThrValTrpMetAlaTyrAspThrGly----- 765
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2039 GGTGCGGTTGG---GCTTATGATTACGGTACAGCAGACAGCAAGAGAAATGTTTGA 2095

QY 766 TyrThrCluArgThrMetAspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySer 785
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2096 TACCTGAAGTCTTATTCTCCGGTA-----CATACGTAAAAAGCGGAAC 2140

```

Search completed: December 12, 2002, 13:55:08
Job time : 103 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 12, 2002, 12:09:28 ; Search time 1947 Seconds

(without alignments)
7178.385 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646

Sequence: 1 MATTGTPTRDGRDAATDPD.....CPESGHEVTLHLPIQEVYL 863

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Fgapop 10.0 , Fgapext 0.5

Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer_p2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US09976674/runatc_04122002_162531_7145/app-query.fasta.1.1031
-DB=EST -GFWT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pio -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09976674.ecgn_1.1.1391.etruncat_04122002_162531_7145 -MCPU=6 -ICPU=3
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -WAIT -LONGLOG -DEV_TIMEOUT=120
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST :
1: em_estba :
2: em_esthum :
3: em_estlin :
4: em_estmu :
5: em_estov :
6: em_estpl :
7: em_estro :
8: em_hic :
9: gb_est1 :
10: gb_est2 :
11: gb_hic :
12: gb_est3 :
13: gb_est4 :
14: gb_est5 :
15: em_estfun :
16: em_estom :
17: gb_gss :
18: em_gss_hum :
19: em_gss_liv :
20: em_gss_pln :
21: em_gss_vtl :
22: em_gss_fun :
23: em_gss_man :
24: em_gss_mus :
25: em_gss_other :
26: em_gss_pio :
27: em_gss_rtd :
27: em_gss_rtd :

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2887.5	62.2	3143	11 AK016546	AK016546 Mus muscu
2	1638	35.3	1137	11 BM553230	BM553230 AGENCOURT
3	1602	34.5	1043	13 B0051605	B0051605 AGENCOURT
4	1516	32.6	921	14 B0877413	B0877413 AGENCOURT
5	1512.5	32.6	932	14 B0684956	B0684956 AGENCOURT
6	1486.5	32.0	982	14 B0678015	B0678015 AGENCOURT
7	1468	31.6	880	14 B0897707	B0897707 AGENCOURT
8	1441	31.0	879	14 B0895618	B0895618 AGENCOURT
9	1439	31.0	879	14 B0889360	B0889360 AGENCOURT
10	1432	30.8	866	13 BM461814	BM461814 AGENCOURT
11	1410	30.3	872	14 B0689588	B0689588 AGENCOURT
12	1403	30.2	910	14 B0949519	B0949519 AGENCOURT
13	1395	30.0	1051	14 BM915935	BM915935 AGENCOURT
14	1383	29.8	821	13 B1690774	B1690774 603314392
15	1372.5	29.5	902	13 B0642814	B0642814 AGENCOURT
16	1369	29.5	787	13 B1233668	B1233668 602943581
17	1360.5	29.3	938	13 B0685669	B0685669 AGENCOURT
18	1320.5	28.4	1030	13 BM424091	BM424091 AGENCOURT
19	1308	28.2	1195	13 BM556123	BM556123 AGENCOURT
20	1292.5	27.8	974	13 BM423964	BM423964 AGENCOURT
21	1274.5	27.4	975	14 B0681307	B0681307 AGENCOURT
22	1272	27.4	1223	14 BM809492	BM809492 AGENCOURT
23	1269.5	27.3	896	13 BG921463	BG921463 602824365
24	1246	26.8	747	13 B1258117	B1258117 602970550
25	1226.5	26.4	997	14 B0642151	B0642151 AGENCOURT
26	1207	26.0	880	14 B0681942	B0681942 AGENCOURT
27	1195	25.7	822	12 BG330712	BG330712 602430226
28	1194.5	25.7	785	13 B1148072	B1148072 602912419
29	1188	25.6	727	13 B1855677	B1855677 603383205
30	1177	25.3	698	12 BG108176	BG108176 602280042
31	1169	25.2	1032	12 BG167854	BG167854 602339821
32	1165.5	25.1	1023	14 B0681703	B0681703 AGENCOURT
33	1164	25.1	809	13 BM045258	BM045258 603623210
34	1154.5	24.8	1080	12 BF203783	BF203783 601868764
35	1153	24.8	1042	13 BM557438	BM557438 AGENCOURT
36	1133	24.4	908	12 BG767446	BG767446 602741246
37	1129.5	24.3	828	12 BG424090	BG424090 602448161
38	1123	24.2	725	13 BM018261	BM018261 603645911
39	1115	24.0	881	12 BF535546	BF535546 602050541
40	1110	23.9	668	12 BG290429	BG290429 602388277
41	1094.5	23.6	1077	13 BM560972	BM560972 AGENCOURT
42	1094	23.5	892	13 B1160407	B1160407 602864430
43	1089	23.4	798	12 BG178503	BG178503 602328312
44	1081	23.3	931	12 BG421210	BG421210 602451701
45	1078	23.2	1088	13 BM458946	BM458946 AGENCOURT

ALIGNMENTS

RESULT 1
AK016546
LOCUS
DEFINITION AK016546 3143 bp mRNA linear HTC 19-JAN-2002
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:932434F09;homolog to DIPEPTIDYL PEPTIDASE 8, full
insert sequence.
ACCESSION AK016546
VERSION AK016546.1 GI:12855334
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (Strain: C57BL/6J) adult male testis cDNA to mRNA,
clone_lib:RIKEN full-length enriched mouse cDNA library
clone:4932434F09.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.


```

Db 462 TATGTCGAACGGTATTCTTGGATGACGTGAAAGAGCTGCTGATACAGAAATATC 521
Qy 45 SerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGlnLysThrAspL 64
Db 522 CACGGCTACATGAATGAGGCGGACATCATCTTATGTTGTAAGAGCGCATCA 581
Qy 65 SerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGlySerArgLys 84
Db 582 GATAGGCTCCTCAAGACAGGCTCATTTACCTTGCATGCTGCGAGACAGAGAAAT 641
Qy 85 SerLeuLeuTyrSerGlyIleProLysLysValArgLysGlnAlaLeuLeuLeuSer 104
Db 642 ACACATATTTATCTGAATCCCTAAACCATCAACAGACAGACAGCTTATGCTTCC 701
Qy 105 TrpLysGlnMetLeuAsnHisPheGlnAlaThrProHisHisGlyValTyrSerArgL 124
Db 702 TGGAAAGCCCTTTGGATCTTTTCAGGCGACACTAGACTATGGGATGATCTCGAGAG 761
Qy 125 GluGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThrSerTyrAspP 144
Db 762 GAAGAGCTACTCAGAGAAAGAGCGCATTTGGAACCGTGGGATCGCAGCTTATGATAC 821
Qy 145 HisSerGlySerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPheHisCysArgAs 164
Db 822 CACCAGAGAACGGCAACTTCCTGTTTCACACTGCGACGGCATTTACCATATCAAGAT 881
Qy 165 GlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIleLysThrGlnC 184
Db 882 GGAGAGCCACATGATTTTACACACAGCCTTGGCGCCAAATTTAGTGAAGACTGTGT 941
Qy 185 SerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPheI 204
Db 942 CCAATATATACAAATGAGATCAAAATATATGCTGCTGATCCAGACTGATAGCTTCAAT 1001
Qy 205 AsnAsnSerAspLeuTyrValAlaAsnIleGluThrGlyGluGluArgLeuThrPhe 224
Db 1002 CACAGCATGATATTTGATATCAAACTCGTAACCGAGAGAGGAGGATCATCAATAC 1061
Qy 225 CysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGlyValAlaThrP 244
Db 1062 GTACACATGAGTACGACCAACATGAGAGATCCACATGACGCGGTGCCACTT 1121
Qy 245 ValIleGlnGluGluPheAspArgPheThrGlyTyrTyrProCysProThrAlaSerTr 264
Db 1122 GTCTCTCAGAGATTTGACAGATCTGCTGCTGCTGCTGCCAACGGAGAAAGA 1181
Qy 265 GluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluValAspLysSerG 284
Db 1182 ACTCCTACTGGTGT---AAATCTTGAATTTCTCTATGAAGAAATGATGATCTGAG 1238
Qy 285 ValGluValIleHisValProSerProAlaLeuGluGluArgLysThrAspSerTyrArg 304
Db 1239 GTGGGATATATTCATGTTACGTCCTCCCATGTTGTAACAGAGGCGGATATCTTCT 1298
Qy 305 TyrProArgThrGlySerLysAsnProLysIleAlaLeuLysIleAlaGluPheGlnThr 324
Db 1299 TATCCCAAAACAGGACGCAAAACCAAGGCTCACTTCAAGATGTGCGAGATTGTTGT 1358
Qy 325 AspSerGlnGlyLysIleValSerThrGlnGlnLysGluLeuValGlnProPheSerSer 344
Db 1359 GATGCTGACGAGGATATATGATGTATGATTAAGAAAGTGTTCACACCTTTCAGATT 1418
Qy 345 LeuPheProLysValGluTyrIleAlaArgAlaGlyTyrPthrArgAspGlyLysTyrAla 364
Db 1419 CTGTTTGGAGGATGAATATATTCAGAGCCGATGACCTCCAGAGGAAACATGCC 1478
Qy 365 TrpAlaMetPheLeuAsnArgProGlnGlnIlePheGlnLeuValLeuLeuProProAla 384
Db 1479 TGGTTCATCTACTAGACGCTTCCAGACTACCTGCGAGATATGTTCTGCTCCCTGAG 1538
Qy 385 LeuPheIleProSerThrGluAsnGluGluArgLeuAlaSerAlaArgAlaValPro 404
Db 1539 TTATTCATCCAGTAGAAGATGATGCCATGACAGACAGAGACTTATAGAGTACTCT 1598
Qy 405 ArgAsnValGlnProTyrValValTyrGluGluValIleThrAsnValTyrIleAsnValHis 424
Db 1599 GACTGTGACACCACTGATCATCTATGAGAAACACAGACATCTGATTAATATCCAC 1658
Qy 425 AspIlePheTyrProPheProGlnSerGluGlyLysAspGluLeuCysPheLeuArgAla 444
Db 1659 GATATTTTTCATGTTTTCCTCAAACTCAT---GAAGATGAATATGATTTATTTTCC 1715
Qy 445 AsnGluCysLysThrGlyPheCysHisLeuTyrLysValIleAlaValLeuLysSerGln 464
Db 1716 TCTGATTCACAAACAGGTTTTCGTCATCTGTATTAATACATCATTTTAAAGAGAGC 1775
Qy 465 GlyTyrAspTrpSerGluProPheSerProGlyGluAspGluPheLysCysProIleLys 484
Db 1776 AAATATTAACGGTCCAGTGGTGAGTACTGCTGCCCAAGTGAATTCATATGCTATCAA 1835
Qy 485 GluGluIleAlaLeuThrSerGlyGluTyrGluValIleAlaArgHisGlySerLysIle 504
Db 1836 GAAGAAATACAAATTAACAGTGTGATGATGGAAGTACTTTGGCCGATGATTAATATC 1895
Qy 505 TrpValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAspThrProLeuGlu 524
Db 1896 TGGGTGATGAAGCCAGAAAGCTGGTACTTGAAGGACACCAAGACTCTCTTGGAG 1955
Qy 525 HisHisLeuTyrValValSerTyrGlnAlaAlaGlyGluIleValArgLeuThrThrPro 544
Db 1956 CATCACTGTACGTGACGACGATTCACAAACCTGGAGAGAGTGGTGCAGCTGATCACC 2015
Qy 545 GlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheValSerHisTyrSer 564
Db 2016 GGCTACTACACTCTGCTGCTGCTGACCGGCACTTGTGACTTCTTATAGTAGTACAGC 2075
Qy 565 SerValSerThrProProCysValHisValTyrLysLeuSerGlyProAspAspAspPro 584
Db 2076 AACAGAGAAATTCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2135
Qy 585 LeuHisLysGlnProArgPheThrAlaSerMetGluAlaAlaSerCysProProAsp 604
Db 2136 GTTCATTAACAAACAGAAATTTTGGGCCACCATTTTGGATTCAGCAGGCTCTCTG 2195
Qy 605 TyrValProProGlnIlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMet 624
Db 2196 TACACCTCTCGAATATTTTCTTTGAAATACTGATGATTTTACACTGATGGAATG 2255
Qy 625 IleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyr 644
Db 2256 TTGTTAAGCCCTACGACTGATCAACCTGGAAGAAATACCCAGCTGTTATTCATATAT 2315
Qy 645 GlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeuArgLeu 664
Db 2316 GGTGTCTCCACAGTCCACTGCTGTAACATCGTTTAAAGAGTCAATATATTTCCGCTG 2375
Qy 665 AsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySerCysGln 684
Db 2376 AACACCTGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2435
Qy 685 ArgGlyLeuArgPheGluGluValAlaLeuLysAsnGlnMetGlyGlnValGluIleGluAsp 704
Db 2436 CGAGAGCTTAATTTGAAGGCGCTTTAAATTAATTAATTAATTAATTAATTAATTAAT 2495
Qy 705 GlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLysSerArgVal 724
Db 2496 CAAGTGAAGGACTCAGTACCTACATCTCAGTATGATGATGATGATGATGATGATG 2555
Qy 725 AlaIleHisGlyTyrSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisTyrPro 744
Db 2556 GGCATTCACAGGCTGCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2615
Qy 745 GlnValPheLysValAlaIleAlaGlyAlaProValIleThrValTyrMetAlaTyrAspThr 764
Db 2616 GATATCTTCCGGGTGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2675

```


(Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

BASE COUNT 209 a 265 c 269 g 188 t 1 others

Alignment Scores:

Pred. No.: 1,81e-170 Length: 932
 Score: 1512.50 Matches: 295
 Percent Similarity: 96.45% Conservative: 4
 Best Local Similarity: 95.16% Mismatches: 9
 Query Match: 32.55% Indels: 4
 DB: 14 Gaps: 1

US-09-976-674-3 (1-863) x B0684956 (1-932)

QY 199 AlapheserPhelieasnsSersaspieutryValaAlaansilegThGlyclu 218
 Db 1 GCGTTCCTTCCTTCATCAATACACGACCTGTGGGTGCAACATCGACAGCGGAG 60
 QY 219 GluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProlySer 238
 Db 61 GAGGCGGCGCTGACCTTCGACCAAGTTTATCCAAATGCTGGATGACCCCAAGTCT 120
 QY 239 AlaglyValAlaThrPheValileGlnGluUpheAspArgPheTrpGlyTrpTrp 258
 Db 121 GCGGGTGGGCGACCTTCATCATACAGAGAGTTCCAGCCCTCAGCGGTACTGTGTG 180
 QY 259 CysProThrAlaSerTrpGlyGlySerGlnGlyLeuLysThrLeuArgileLeuTrgln 278
 Db 181 TGCCCGACAGCTCTCTGGAGGTTTCAGAGGCGCTCAAGACGTGCGAATCTGTATGAG 240
 QY 279 GluValAspGluSerGluValileHisValProSerProAlaLeuGluGluArg 298
 Db 241 GAACTCATGAGTCCGAGGTGAGGTATTCACCTCCCTCTCTGCGTAGAAGAAAG 300
 QY 299 LysThrAspSerTrpArgTrpProArgThrGlySerLysAsnProLysileAlaLeuLys 318
 Db 301 AAGACGAGCTCGATCGGTACCCAGAGGACAGCAAGATCCCAAGATTGCTTGAAA 360
 QY 319 LeuAlaGluPheGlnThrAspSerGlnGlyLysileValSerThrGlnGluLysGluLeu 338
 Db 361 CTGGCTAGTTCACAGCTGACAGCGGCGCAAGATCTCTGACCCAGAGAAAGAGCTG 420
 QY 339 ValGlnProPheSerSerLeuPheProLysValGluTrpIleAlaArgAlaGlyTrpThr 358
 Db 421 GTGAGCGCTTCACCTGCTGTTCCGAGAGGTGAGTACATGCGCAGGCGGCTGAGCC 480
 QY 359 ArgAspGlyLysTrpAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeu 378
 Db 481 CGGATGGAATACGCGCTGGCGCATGCTCTGAGACCGCCCAAGTGGCTCCAGCTC 540
 QY 379 ValLeuLeuProProAlaLeuPheIleProSerThrGluAsnGlnGluGlnArgLeuAla 398
 Db 541 GTCTCTCTCCCGCGCGCTGTTCATCCGACAGAGATGAGAGACAGCGCTAGCC 600
 QY 399 SerIleArgAlaValProArgAsnValGlnProTrpValValTrpGluGluValThrAsn 418
 Db 601 TCTGCCAGAGCTGTCGCCAGGAATGTCAGCGATGCTGCTGAGAGAGAGTACCAAC 660
 QY 419 ValTrpIleAsnValHisAspIlePheTrpProPheProGlnSerGlu-GlyGluAspG1 438
 Db 661 GTCGAGATCAATGTTTCATGACATCTTCATCCCTCCCAATCAGAAAGAGAGAGAGA 720
 QY 438 uLeuCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTrpValTrh 458
 Db 721 GCTCTGCTTCTCCGCCCATGATGATCAACACGCGCTTCGCAATTGTCTCAAAAGTCAC 780
 QY 458 rAlaValLeuLysSerGlnGlyTrpAspTrpSerGluProPheSerProGlyGluAspG1 478
 Db 781 CGCCGTTTAAAT-TCCAAGGGCTACCAATTGAGTGA-CCCTTGCGCCCGGGGAGAGAGA 838
 QY 478 uPheLysCysProIleLysGlnGluIleAlaLeuThrSerGlyGluTrp---GluValle 497
 |||||||

Db 839 ATTAAGTCCCATTAAGAGAGATGCTCTGACCAAGCGGGAATAATGGGAGGTTT 898
 QY 497 uAlaArgHisGlySerLysIleTrpVal 506
 Db 899 GCGGAGCAGCGCTCAAAATTTGGGTT 926

RESULT 6
 LOCUS B0678015
 DEFINITION B0678015 982 bp mRNA linear EST 15-JUL-2002
 AGENCOURT_8034317 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6213986
 5', mRNA sequence.
 ACCESSION B0678015
 VERSION B0678015.1 GI:21790694
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: DCTD/DRP

cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LICM2380 row: 3 column: 03
 High quality sequence stop: 683.

FEATURES
 source 1..982
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6213986"
 /clone_id="NIH_MGC_112"
 /tissue_type="melanotic melanoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site: 1; XhoI; Site: 2; EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
 BASE COUNT 202 a 310 c 290 g 179 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 2.69e-167 Length: 982
 Score: 1486.50 Matches: 295
 Percent Similarity: 93.40% Conservative: 2
 Best Local Similarity: 92.77% Mismatches: 17
 Query Match: 32.00% Indels: 6
 DB: 14 Gaps: 1

US-09-976-674-3 (1-863) x B0678015 (1-982)

QY 489 LeuThrSerGlyGluTrpGluValLeuAlaArgHisGlySerLysIleTrpValAsnGlu 508
 Db 3 CTGACCAAGCGGTGATGGAGGTTTGGCGAGCAGCGCTCAAGATCTGGTCAATGAG 62
 QY 509 GluThrLysLeuValTrpPheGlnGlyThrLysAspTrpProLeuGlnHisLeuTrp 528
 Db 63 GAGACCAAGCGGTGATCTCCAGGGCACCAAGACAGCGCGGTGAGACACACTCTAC 122
 QY 529 ValValSerTrpGluAlaAlaGlyGluIleValArgLeuThrTrpProGlyPheSerHis 548
 Db 123 GTGGTACGATGAGGGCGGCGGAGATGCTAGCGCTCACACAGCGCGGCTTCCCAT 182
 |||||||

Oy	549	SerCySerMetSerGlnAsnPhaeSmphePheValSerHisTySerSerValSerThr	568
Db	183	AGCTCTCCATGAGCCAGAACTTCCACATGTTCTGCAGCCACTACAGCAGCGTGAGCACG	242
Oy	569	ProProCyValHisValTyrTyrLeuSerGlyProAspAspProLeuHisLysGln	588
Db	243	CCGCGCTCGTGCACGTGTACAAAGCTGAGCGGCCCGCAGACGACCCCTCGACAAAGCAG	302
Oy	589	ProAspPhePhePheValSerMetMetGlnAlaAspCysProProAspTyrValProPro	608
Db	303	CCCCCTCTTGCGCTAGGATGATGGAGGACGACACTCGCCCCCGGATTATGTTCTCTCA	362
Oy	609	GluIlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysPro	628
Db	363	GAGATCTTCATTTCCACACGCGCTCGGATGTGGCGCTCTACGGGATGATGTACAAAGCC	422
Oy	629	HisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyrGlyGlyProGln	648
Db	423	CACGCTCTGCACCCGAGGAGAGACCCACCGCTCTTGTGTATGTAGAGGCCCCAG	482
Oy	649	ValGlnLeuValAsnAspSerPheLysGlyIleLysTyrLeuArgLeuAsnThrLeuAla	668
Db	483	GTCGACGTGTGTATTAATCTCTTCAAAGGCATCAAGTACTTCCGGCTCAACACACTGGCC	542
Oy	669	SerLeuGlyTyrTyrAlaValValAlaLeaspGlyArgGlySerCysGlnArgGlyLeuArg	688
Db	543	TCCCTGGGCTACCGCGTGTGTGTGATGTAGTCAGGCAAGGCGCTCTGTACAGAGGCTTGG	602
Oy	689	PheGluGlyAlaLeuLeuLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGly	708
Db	603	TTTCGAGGAGGCCCTTAATAAACCAAAATGGGCCCGGTGAGATCGAGGACACGAGTGAAGGC	662
Oy	709	LeuGlnPheValAlaGluLysTyrGlyPheLeuAspLeuSerArgValAlaIleHisGly	728
Db	663	CTGCAGTTCGTGGCGGAGAGATAGGCTTCATCGACACTGAGCGAGTTGCCATTCATGGC	722
Oy	729	TrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLysPheProGlnValPheLys	748
Db	723	TGGTCTCTACGGGGGCTCTCTCTCGCTCATGGGGCTTAATATCC-AACCCCAAGTGTCTCAAG	781
Oy	749	ValAlaIleAlaGlyAlaA-ProValThrValTyrMetAlaTyrAspThrGlyTyrThrGly	768
Db	782	GGGGCATTCGGGGGCGCCCGCGCTACCGCTGTGATGGCTCTACGACACAGGTTACACTGA	841
Oy	768	uArg-TyrMetAspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaL	788
Db	842	GCGGCTACATGGAACGTCCTGTGAAACAACAG-CACGGCTTAGAAGAGGGGGTTCGGGGCC	900
Oy	788	eunHisVal---GluLysLeu-ProAsnGluProAsnArgLeuLeuIle 802	
Db	901	CCTGGACGGGGGAAAACTGCCCCCAGGAAACCAACCCGCGCTTGTCTT 948	
RESULT 7			
LOCUS	B0897707	880 bp	mRNA linear EST 16-AUG-2002
DEFINITION	AGNCNCURT.8061873 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208407		
ACCESSION	B0897707		
VERSION	B0897707.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	NIH-MGC http://mgs.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cga@bbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Rubin Laboratory		

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
DNA Sequencing by: Agencourt Bioscience Corporation				
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov				
Plate: L1CM2366 row: a column: 16				
High quality sequence stop: 704.				
Location/Qualifiers				
source	1.	880		
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:6208407"			
	/clone_1lb="NIH_MGC_110"			
	/tissue_type="ductal carcinoma, cell line"			
	/lab_host="DH10B (phage-resistant)"			
	/note="Organ: pancreas; Vector: pCMV7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(g). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies Note: this is a NIH_MGC library."			
BASE COUNT	194 a	260 c	249 g	176 t 1 others
ORIGIN				
Alignment Scores:				
Pred. NO.:	3,69e-165	length:	880	
Score:	1468.00	Matches:	287	
Percent Similarity:	97.95%	Conservative:	0	
Best Local Similarity:	97.95%	Mismatches:	0	
Query Match:	31.60%	Indels:	6	
DB:	14	Gaps:	0	
US-09-976-674-3 (1-863) x BQ897707 (1-880)				
OY	139	AlAPhPheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlu	218	
Db	1	GCGTCTCTCTCTTCATCAATACAGAGCACTGGGGTGCCCAATCGAGACAGCGGAG	60	
OY	219	GIuArGArLeuThrPheCysHISG1G1yLeuSerzsnValLeuAspAPrPolysSer	238	
Db	61	GAGCGGGCGGTGACCTTGCCACCAAGGTTATCCAATGCTCGATGAGACCCCAAGTC	120	
OY	239	AlaGlyValAlaThrPheValIleGlnGluIuPheAspArgPheThrGlyTrpTrp	258	
Db	121	GCGGGTGGCCACCTTCGTATCAGAAAGTTCACGCCCTTCCTGGGTACTGGTGG	180	
OY	259	CysProThrAlaSerTrpGluG1ySerGluIyLeuysTrhLeuAgiLeuTrpGlu	278	
Db	181	TGCCCAACAGCCTCTCGGGAAGGTTCCAGAGGGCTCAAGACGCTGGCAATCTGTATGAG	240	
OY	279	GIuValAspGluSerGluValGluValIleHisValProSerProAlaLeuGluIuArg	298	
Db	241	GAAGTCGATGATGCCAGAGGTGAGAGTCATTCACCTCCCTCTCTCGGCTAGAAAGG	300	
OY	299	LysThrAspSerTrpArgTrpProArgThrGlySerLysAsnPolysIleAlaLeuLys	318	
Db	301	AAGAGGCACTGTATCGGTATCCCAAGGACGGCAGCAAGATCCCAAGATTCCTTGAAA	360	
OY	319	LeuAlaGluPheGlnThrAspSerGlnG1yLysIleValSerThrGlnGluIySgIuLeu	338	
Db	361	CTGGCTGAGTTCAGAGCTGAGCAAGCCAGCAAGATGCTGTCTGCAGCCAGGAAAGAGCTG	420	
OY	339	ValGlnProPheSerSerLeuPheProLysValGluTrpIleAlaArgAlaGlyTrpThr	358	
Db	421	GTCGAGCCCTTCAGCTGCTGTTCCCAAGGTGAGTACATCGCCAGCGCGGCTGGAC	480	
OY	359	ArgAspGlyLysTrpAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeu	378	
Db	481	CGGATGGCAATACGCTGGGCGCATGTTCTTGAGACGGGCCCAAGCATGTGCTCCAGCTC	540	
OY	379	ValLeuLeuProProAlaLeuPheIleProSerThrGluAsnGluGlnArgLeuAla	398	

Dn	541	GTCCCTCCCCCGGCGCCTTTCATCCCGAGCACAGAAATGAGACAGGGCTAACC	600
Oy	399	SeralaaagalaavalproargasnaValglInProTyrrValValTYrGlUGluValThrAsn	418
Dn	601	TCTGCACAGAGCTGTGCCAGGAATGTCCAGCGCATGTGCTGTACGAGAGGTCAACAAC	660
Oy	419	VAlTrPIleAsnVAlHISASPlIePherTyPrOPhePrOGlnSerGIUGluASPlU	438
Dn	661	GCTGGAACCAATGTGTCAACACATCTTCATCCCTTCCCACATGAGAGGGAGAGAGAG	720
Oy	439	LeucysPHeIeuArGaLaSnGLUCysLysThrgILyPHeCySHSLeuTyrrLyValThr	458
Dn	721	CTTGCTTTCTTCGCCGCAATGATGACAGACGGGCTTCTGCATTGTGTACAAGAATCAC	780
Oy	459	Ala-ValleuLysSergIn-glyTyR-ASPTrPSergIU-ProPheSer-ProGLyGUa	477
Dn	781	GCCCGTTTTAAATGCCAGGGGGCTACGCAATTGGAGTAGACGCCCTTACACCCCGGGGAG	840
Oy	477	sPGluPheLysCyS-ProIIeLysGlu	485
Dn	841	ATGAATTTAAGTGGCCCCCATTAAGGAA	867
RESULT 8			
B0895618			
LOCUS		943 bp	mRNA linear EST 16-AUG-2002
DEFINITION	AGENCOURT_8753137 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6394542		
ACCESSION	5'', mRNA sequence.		
VERSION	B0895618		
KEYWORDS	B0895618.1 GI:22287632		
SOURCE	EST.		
ORGANISM	house mouse.		
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/.		
AUTHORS	(bases 1 to 943)		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs@email.nih.gov Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D. cDNA Library Preparation: ResGen, Invitrogen Corp cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLML13887 row: m column: 07 High quality sequence stop: 729. Location/Qualifiers		
FEATURES			
source	1..943		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:6394542"		
	/clone_id="NIH_MGC_130"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: otcocysts; Vector: pCMV-SPORT6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.95 kb. Constructed by ResGen, Invitrogen Corp. Note: This is a NIH_MGC Library."		
BASE COUNT	216 a 276 c 256 g 193 t	2 others	
ORIGIN			
Alignment Scores:			
Pred. No.:	7,48e-162	Length:	943
Score:	1441.00	Matches:	260
Percent Similarity:	93.07%	Conservative:	22
Best Local Similarity:	85.81%	Mismatches:	20
Query Match:	31.02%	Indels:	1
Dn:	14	Gaps:	0

[illegible]

VERSION BQ889360.1 GI:22281374
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2366 row: g column: 16
High quality sequence stop: 709.
Location/Qualifiers
1. 879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:620851"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 194 a 256 c 252 g 177 t
ORIGIN
Alignment Scores:
Pred. No.: 1.14e-161 Length: 879
Score: 1439.00 Matches: 279
Percent Similarity: 96.58% Conservative: 3
Best Local Similarity: 95.55% Mismatches: 5
Query Match: 30.97% Indels: 5
DB: 14 Gaps: 2
US-09-976-674-3 (1-863) x BQ889360 (1-879)
QY 199 AAlaphePheSerPheIleAsnAsuSerAspLeuTrpValAlaAsnIleGluThrGlyGlu 218
|||||
Db 1 GCCTCTTCCTTCATCAATACAGGACCTGTGGGTGGCCAAATCGAGACGGGGAG 60
QY 219 GluAtrGrArgLeuThrPheCySHsGlnGlyLeuSerAsnValLeuAspAspProLysSer 238
|||||
Db 61 GACGGGGGCTGACCTCTCCACCAAGGTTTATCCAAATGCTCCGATGACCCCAAGTCT 120
QY 239 AAlGlyValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrp 258
|||||
Db 121 GCGGGGTGTGGCCACCTTCATACAGGAAGAGTTCACCGCTTCACTGGGTACTGTGG 180
QY 259 CysProThrAlaSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlu 278
|||||
Db 181 TGCCCCCAGACCTCTCTGGGAAGGTTCAGAGGCTCAAGACGCTGCAATCTGTATGAG 240
QY 279 GluValAspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArg 298
|||||
Db 241 GAAGTCATAGATCCGAGGTGAGAGTCATTCACGCTCCCTCTCGGTGAGAAAGAG 300
QY 299 LysThrAspSerTyrArgTyrProArgTyrGlySerLysAsnProLysIleAlaLeuLys 318
|||||
Db 301 AAGACGACTCGTATCGTACCCAGACAGGACAGCAAGATCCCAAGATTGGCTTGAAA 360

QY 319 LeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeu 338
|||||
Db 361 CTGGCTAGTTCAGACTGACAGGACGCAAGATCTCTCGACCCAGGAAGAGACTG 420
QY 339 ValGlnProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrPThr 358
|||||
Db 421 GTGAGAGCCCTTACGCTGCTGTCCCAAGGTGAGATCATCCAGGCGCGGTGGACC 480
QY 359 ArgAspGlyLysTyrAlaTATPAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeu 378
|||||
Db 481 CGGATGGCAAAATACGCTGGCCATGCTCTGGACCGGCCCAAGCAGTGGCTCAGCTC 540
QY 379 ValLeuLeuProProAlaLeuAlaLeuPheProSerThrGlnAsnGlnGluGlnArgLeuAla 398
|||||
Db 541 GTCTCTCTCCCGCGCGGCGGCTTCATCCGACGACAGAGAATAGAGAGCGGCTAGCC 600
QY 399 SerAlaArgAlaValProArgAsnValGlnProTyrValValTyrGluValThrAsn 418
|||||
Db 601 TCTGCCAGACCTCTCCCGACGAATGTCACCGCATGTGTGTACGAGGAGTCCACCAC 660
QY 419 ValTrpIleAsnValHisAspIlePheTyrProPheProGlnSerGluGluAspGlu 438
|||||
Db 661 GTCTGATCATGTTCATGACATCTTATCCCTTCCCAATCAGAGGAGAGAGAG 720
QY 439 LeuCySPheLeuArGAlaAsnGluCySPheGlyPheCySHsLeuTyrLysValThr 458
|||||
Db 721 CTCTGCTTCTCGCGCGCATGATGACGACGCGGCTTCGCCATTGTGTCAAAGTAC 780
QY 459 Ala-ValLeuLysSerGln-GlyTyrAsp-TrpSerGlnProPheSer---ProGlyGlu 476
|||||
Db 781 GCCCTTTTAAATCCAGGCGGTACGATTTGGATGAGTACGCCCTTCAGCCCGGGGAG 840
QY 477 AspGluPhe---LysCySProlIleLysGlu 485
|||||
Db 841 GATGAATTTTAAGTGGCCCATTAAGAA 870
RESULT 10
BM461814
LOCUS BM461814
DEFINITION 866 bp mRNA linear EST 05-FEB-2002
5', mRNA sequence.
ACCESSION BM461814
VERSION BM461814.1 GI:18510854
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM12219 row: g column: 20
High quality sequence stop: 627.
Location/Qualifiers
1. 866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:5533867"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb. "

BASE COUNT 188 a 269 c 245 g 164 t

ORIGIN

Alignment Scores:

Pred. No.: 7,72e-161 Length: 866
 Score: 1432.00 Matches: 276
 Percent Similarity: 97.87% Conservative: 0
 Best Local Similarity: 97.87% Mismatches: 2
 Query Match: 30.82% Indels: 4
 DB: 13 Gaps: 0

US-09-976-674-3 (1-863) x BM461814 (1-866)

QY 40 Glycerarglytyrserglyleuilevalasnlalsalaprohisaspheglnpheval 59
 |||||
 Db 11 GGCAGCGGCAAGTCTGGGCTCATGTCTCAACAGCGCCGCCAGACTTCCACTTGTG 70
 QY 60 Glnythrapsrgluserglyprohisserhisargleutytyrleuglymetprotyr 79
 |||||
 Db 71 CAGAGACGATGATGTGGGCCCTCCACCGCTCTACTACTGGAATGCCATAT 130
 QY 80 Glyserargluserleuileutysergluileprolyslvalargylsluala 99
 |||||
 Db 131 GGCAGCGGAGAACCTCCCTCTACTCTGAGATCCCAAGAGGTCCGGAAGAGGCT 190
 QY 100 Leuileuileuusertrpilyslmetleuaspheglnalathrprohisglly 119
 |||||
 Db 191 CTGTGTCTCTCTCTCTGGAAGCATCTGTGATCATTTCCAGGCGCCACCATATGG 250
 QY 120 Valtyrserarggluglulgluileuarglualargylsargyleuglyvalphegly 139
 |||||
 Db 251 GTCTACTCTGGGAGGAGAGCTCTGAGGAGGGAACCCCTGGGGGCTTGGCATC 310
 QY 140 Thisertryasphehissergluserglyleupheleupheglnalaserasnerleu 159
 |||||
 Db 311 ACCGCTACGACTTCCACAGGAGAGTGGCTCTCTCTCCAGGCGCAGCAAGCCTC 370
 QY 160 Phehisysaragspilyglylyslasnglyphenetvalserprometlysproleuglu 179
 |||||
 Db 371 TTCCACTGCGCGGCGGCGGCAAGACGGCTTCATGTGTCCTATGAACCCCTGGA 430
 QY 180 Ilelythrnglncyserserglyproargmetasprolysliecyssproalasprrala 199
 |||||
 Db 431 ATCAAGACCAAGTCTCAAGGCCCCGATGACCCCAAAATCTCCCTGCGGACCTGCC 490
 QY 200 Phepserpheleasnasnsersaspheutryvalalaasnilegluthrglyglu 219
 |||||
 Db 491 TTCTTCTCTCATCATATACAGGAGCTGTGGTGGCCAAATCGAGACAGGCGAGAG 550
 QY 220 Argargleuthrphecysnhsinglyleuserasnavalleuasasprralysserala 239
 |||||
 Db 551 CGGGGCTGACCTTCTCCACCAAGGTTTATCCAAATGTCTCGATGACCCCAACTGCG 610
 QY 240 Glyvalalathrphavalileglnlglupheaspargpethrglytyrtrprrcys 259
 |||||
 Db 611 GGTGTGGCCACTTCATATACAGAGAGTTGACCGCTTCACTGGGTACTGTGTGGC 670
 QY 260 Prothrthlasertprgluglysergluglyleutystrhleuargileutyrglu 279
 |||||
 Db 671 CCCACACCTCTCTGGAGGGTCAAGAGGCTCAAGACGCGGCAATCTGTATAGAGAA 730
 QY 280 Valaspluser-gluvalgluvalilehsvalproserproalaleugluu-argl 299
 |||||
 Db 731 GTGATATAGTCCGAGGTGGGTGATTCAGTCCCTCTCTCTGCTAGAGAAAGAA 790
 QY 299 ystrhpsersertryargtyr-proargthrglyserlys-asnprolysllealatala 317
 |||||
 Db 791 AGACGGACTGTATCGGTACCCCAAGACAGGACGCAAGGATCCCAAGAAATGCGCTT 848

RESULT 11
 B0689588

LOCUS B0689588 872 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT_8345928 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6250957
 5', mRNA sequence.
 ACCESSION B0689588
 VERSION B0689588.1 GI:21814904
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 872)
 NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LNCM394 row: n column: 14
 High quality sequence stop: 654.
 Location/Qualifiers
 source
 1. 872
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6250957"
 /clone_1bp="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGC(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."
 BASE COUNT 192 a 258 c 247 g 174 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 3.48e-158 Length: 872
 Score: 1410.00 Matches: 274
 Percent Similarity: 96.15% Conservative: 1
 Best Local Similarity: 95.80% Mismatches: 6
 Query Match: 30.35% Indels: 5
 DB: 14 Gaps: 2
 US-09-976-674-3 (1-863) x B0689588 (1-872)
 QY 199 Alapheserpheleasnasnsersaspheutryvalalaasnilegluthrglyglu 218
 |||||
 Db 1 GCGTCTTCTCTCATATACAGGAGCTGTGGTGGCCAAATCGAGACAGGCGAG 60
 QY 219 Gluargargleuthrphecysnhsinglyleuserasnavalleuasasprralysser 238
 |||||
 Db 61 GAGCGGGGCGGACTTCTGCGACCAAGGTTATCCATGTCCCGATGACCCCAAGTCT 120
 QY 239 Alaglyvalalathrphavalileglnlglupheaspargpethrglytyrtrprrcys 258
 |||||
 Db 121 GCGGTGTGGCCACTTCGATATACAGAGAGTTCACCGCTTCACTGGGTACTGTGTGG 180
 QY 259 Cysprothrthlasertprgluglysergluglyleutystrhleuargileutyrglu 278
 |||||
 Db 181 TGCCCAAGGCTCTCTGGAGGTTTCAAGGCTCAAGAGGCTGCGAATCTGTATAG 240
 QY 279 Gluvalasplusergluvalgluvalilehsvalproserproalaleugluuarg 298
 |||||
 Db 241 GAATCGATGATGTCGAGGTGGAGGTGATTCACCTCCCTCTCTGCGCTAAGAAAGG 300

QY 299 LysThrAspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLys 318
|||||
Db 301 AAGCGAGACTCGATCGTACGCCAGACAGGACAGCAAGATCCAGATTCGCTCGAAA 360
QY 319 LeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeu 338
|||||
Db 361 CTGGCTGAGTTCAGACTGACAGCAGGCGAAGTCTCTCCAGCCAGGAGAAAGACCTG 420
QY 339 ValGlnProPheSerSerLeuPheProLysValGlnTrpIleAlaArgAlaGlyTrpThr 358
|||||
Db 421 GTGCAGCCCTTCAGCTCGCTGTTCCCGAAGGTGAGTACATCGCCAGGCGCGGTGACCC 480
QY 359 ArgAspGlyLysTyrAlaTrpAlaMetPheLeuAsnParProGlnGlnTrpLeuGlnLeu 378
|||||
Db 481 CGGAGTGCAGAAATACGCGCTGGCCATGTCCTGGACCGGCCAGCAGTGGCTCCAGCTC 540
QY 379 ValLeuLeuProProAlaLeuPheIleProSerThrGluAsnGluGlnAlaArgLeuAla 398
|||||
Db 541 GTCCTCTCCCGCCCGCCGCTTCATCCAGACAGATGAGAGACGCGGCTAGCC 600
QY 399 SerAlaArgAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAsn 418
|||||
Db 601 TCTGCCAGACTGTCCCGAGGAATGTCCAGCCGATGTGTTGACGAGAGGTACCAAC 660
QY 419 ValTrpIleAsnValHisAspIlePheTyrProPheProGlnSerGlnGluAspGlu 438
|||||
Db 661 GTCTGATCAATGTTTCATGACATCTTATCCCTTNCCTCCCAATCAGAGGAGAGACGAG 720
QY 439 LeuCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValIth 458
|||||
Db 721 CTCTGCTTCTCCGCGCCCAATGAAATGCAACGCGCTTCTGCAATTTGACAAAGTCCAC 780
QY 458 AlaValLeuLysSerGlnGlyTyrAspTrpSer---GluProPheSerPro---Gly 475
|||||
Db 781 CGCGCTTTTAAATCCCGACAGGCTACCGATTGGAGTGAAGCCCTTAAGCCCCGGGGA 840
QY 476 GluAspGluPhe 479
::: |||||
Db 841 AAGATGGAATTT 852

RESULT 12
LOCUS BQ949519 910 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCODRT_8794758 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6374913
5', mRNA sequence.
BQ949519
VERSION BQ949519.1 GI:22364997
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 910)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: DCTD/DFP/Gazdar
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2534 row: k column: 10
High quality sequence stop: 668.
Location/Qualifiers
1..910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6374913"

/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="Dh10B (phage-resistant)"
/note="Organ: Lung; Vector: pOTB1; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
ggcgcagag(g). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
SuperScript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 190 a 294 c 261 g 165 t
ORIGIN

Alignment Scores:
Pred. No.: 2,63e-157 Length: 910
Score: 1403.00 Matches: 275
Percent Similarity: 96.17% Conservative: 1
Best Local Similarity: 95.82% Mismatches: 7
Query Match: 30.20% Indels: 5
DB: 14 Gaps: 0

US-09-976-674-3 (1-863) x BQ949519 (1-910)

QY 1 MetaLatThrGlyThrProThrAlaAspArgGlyAspAlaAlaIleThrAspAspPro 20
|||||
Db 37 ATGGCCACACCGGAGCCCAACAGCGCGCAGCGAGCGACGCGCCACAGATGACCCG 96
QY 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
|||||
Db 97 GCGCGCCGCTTCAGAGGCGAAGACCTCGTGACGCGGTCCGCGACATCATCCACGGC 156
QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
|||||
Db 157 AGCCGAGAGACTCGGGCCCTCATGTGTCAACAAGCGGCCCAAGACTCCAGTTGTGTCAG 216
QY 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTrpLeuGlyMetProTyrGly 80
|||||
Db 217 AAGACGATAGTGTGGCCCACTCCACCGCTCTACTACCTGGAATGCCATATGGC 276
QY 81 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
|||||
Db 277 AGCCGAGAGACTCGGGCCCTCATGTGTGTCAACAAGCGGCCCAAGACTCCAGTTGTGTCAG 336
QY 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
|||||
Db 337 CTGCTCTGCTCTGAGAGAGATGCTGTGATCATTTCCAGGCGACGCCCAACATGGGCTC 396
QY 121 TyrSerArgGluGluGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr 140
|||||
Db 397 TACTCTGGGAGGAGGAGCTGTAGGAGCGGAACGCCCTGGGGGTCTTCGGCATCACC 456
QY 141 SerTyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
|||||
Db 457 TCTACACACTTCCACACGAGAGTGGCTCTTCTCTTCCAGGCCACCAACAGCTTTC 516
QY 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
|||||
Db 517 CACTGCCGCGCGCGCGCAAGACGCTTCATGTGTGCCCTTAAGAAACGCGTGAATAC 576
QY 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
|||||
Db 577 AAGACCCAGAGTCTCAGGCGCCCGGATGAGCCCAAAATGTGCCCTCGGACCCCTGCTTC 636
QY 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluTrpArgLysGluArg 220
|||||
Db 637 TTCTCTTCATCAATACACGACGACTGTGGTGCCAACTCGAGACAGGAGACGG 696
QY 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAla-G1 240
|||||
Db 697 CGGCTGACCTTTCGCCCAAA-GGTTTATCCAAATCTCTGGATGACCCCAAGTCTCGCGGG 755
QY 240 yValAlaThrPheValIleGlnGluGluPheAsp-ArgPheThrGlyTyrTrpTrpCysP 260

```

Db      756  TGTGGGACCTTCATACAGAGAGAGTGCAGCCGCTTCACTGGTACTGGTGGCC 815
Qy      260  rothAla-SerTrpGluGlySerGluGlyLeuYsthrLeu-AngIIleuTyrglu 279
Db      816  CCACAGCCCTCTGGGAGGTTTCAGAGAGGCTCTCAGACGCTCGAATCTGTATGTAGA 875
Qy      279  uValaspluSer 283
Db      876  ACTCCATGAGTCC 888

RESULT 13
LOCUS    BM915935
DEFINITION AGNCOURT 6639540 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5482205
5' mRNA sequence.
ACCESSION BM915935
VERSION    BM915935.1 GI:19366314
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1051)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM2007 row: 0 column: 06
            High quality sequence stop: 677.

FEATURES
    source
        1..1051
            location/qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5482205"
                /clone_lib="NIH_MGC_41"
                /tissue_type="amelanotic melanoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: Skin; Vector: pORF7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGCACGAG(g). Library constructed by Ling Hong in the
                laboratory of Gerald M. Rubin (University of California,
                Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                Superscript II RT (Life Technologies). Note: this is a
                NIH-MGC Library."
BASE COUNT 233 a 329 c 300 g 189 t
ORIGIN
Alignment Scores:
Pred. No.: 3.17e-156 Length: 1051
Score: 1395.00 Matches: 280
Percent Similarity: 92.46% Conservative: 2
Best Local Similarity: 91.80% Mismatches: 12
Query Match: 30.03% Indels: 11
Gaps: 3
US-09-976-674-3 (1-863) x BM915935 (1-1051)
Qy      114  AlaThrProHisGlyValTyrSerArgGluGluLeuLeuArgGluArgLysArg 133
Db      2     GGCAGCCCCACACATGGGTTCTACTCTCGGAGAGAGACTGCTGAGGAGCGGAACCC 61
Qy      134  LeuGlyValPheGlyIleThrSerTyrAspPheHisSerGluSerGlyLeuPheLeuPhe 153

```

```

Db      62  CTGGGGGCTCTTCGGCATCACTCTACAGCTTCCACAGGAGAGTGGCTTCTCTTC 121
Qy      154  GluAlaSerAsnSerLeuPheHisCysArgAspGlyGlyAsnGlyPheMetAlaSer 173
Db      122  CAGGCGAGCAACAGCTCTTCCACTGCGCGGAGGGGCAAGAGCGCTTCATGGGTGTC 181
Qy      174  PrometLysProLeuGluIleLeuYsthrGlnCysSerGlyProArgMetAspProLysIle 193
Db      182  CCTATGAACCCGCTGGAATCAAGACCCAGTCTCAGGCCCCGGATGAGACCCCAAAATC 241
Qy      194  CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTrpValAlaAsn 213
Db      242  TCCCTGCGCCGACCTGCTCTTCTTCATCAATAAACAGGACCTGGGGTGCCAAAC 301
Qy      214  IleGluThrGlyGluGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 233
Db      302  ATCGAGACAGGGAGAGGCGGCTGACCTTGCACCAAGGTTTATCCATGCTCCG 361
Qy      234  AspAspProLysSerAlaGlyValAlaThrPheValIleGlnGluGluPheAspArgPhe 253
Db      362  GATGACCCCAAGTCTCGGGGTGCGCCACCTTCGTCATACAGGAAGAGTTCACCGCTTC 421
Qy      254  ThrGlyTyrTrpTrpCysProThrAlaSerTrpGluGlySerGluGlyLeuYsthrLeu 273
Db      422  ACTGGTACTGGTGGTGGCCACAGGCTCTCGGAAGGTTTCAGAGGGCTCAAGACGCTG 481
Qy      274  ArgIIleuTyrgluGluValAspGluSerGluValGluValIleHisValProSerPro 293
Db      482  CCAATCTGTATGAGGAAGTCAATGATGATCCGAGTGAAGTCAATCACTCCCTCTCT 541
Qy      294  AlaLeuGluGluArgLysThrAspSerTyrArgTyrProArgThrGlySerLysAsnPro 313
Db      542  GCGCTAGAAGAAAGAGAGACGACTGTCGTATCGGTACCCAGAGACAGCAAGAAATCC 601
Qy      314  LysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThr 333
Db      602  AAGATTGCTCTTAAGAACTGCTGATTCAGACTGACAGCCAGGCAAGATCGTCTCGACC 661
Qy      334  GlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysValGluIleAla 353
Db      662  CAGGAGAAAGACTGTGTGAGCCCTTCACCTGCTGTTCCAGAAAGTGGGT-----ACT 715
Qy      354  ArgAlaGlyTrp-----ThrArgAspGlyLysTyrAla-TrpAlaMetPheLeuAs 370
Db      716  CCGCAGAGGGGGCGGGGTGGAACCCCGGAATGGCAATATGCGCTGGGCGCATGCTCGA 775
Qy      370  PArg-ProGlnGlnTrpLeuGlnLeuValLeuLeuProPro-AlaLeuPheIleProSer 389
Db      776  CCGGCCCCAGCAGATGGCTCCAGCTGCTCTCTCCGCCCGGGCTGTTCATCCGAGCC 835
Qy      390  ThrGlu--AsnGluGluGlnArgLeu---AlaSerAlaArgAlaValProArgAsnValG 408
Db      836  CCCAAGAATATGAGGAGCAGCGGCTTAACCTCTGCGCAAGTGGCCCAAGAAATGTCC 895
Qy      408  InPro 409
Db      896  AACCC 900

RESULT 14
LOCUS    B1690774
DEFINITION 603314392F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5354182 5',
mRNA sequence.
ACCESSION B1690774
VERSION    B1690774.1 GI:15653403
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 821)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)

```

**JOURNAL
COMMENT**

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1900 row: h column: 23
High quality sequence stop: 821.

FEATURES
source

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:5354182"
/clone_id="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Saliv
Site_2: Noli; Cloned unidirectionally. Primer: Oligo dT
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

```

```
Percent Similarity: 92.26%      Conservative: 5
Best Local Similarity: 90.57%    Mismatches: 11
Query Match: 29.54%             Indels: 12
DB: 14                          Gaps: 2

US-09-976-674-3 (1-863) x BQ642814 (1-902)

QY 282 GluSerGluValGluValIleHisValProSerProAlaLeuGluArgLysThrAsp 301
DB 3 GAGTCCGAGGTGGAGGTCAATTCAGTCCCTCCCTCGCTAGAGAAAGAGACGGAC 62
QY 302 SerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAlaGlu 321
DB 63 TCSTATCGGTACCCAGGACAGCAGCAAGATCCCAAGATTGCCTTGAAACTGGCTGAG 122
QY 322 PheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGlnPro 341
DB 123 TTCAGACTCAGACCCAGGCGAGATCGTCTCGACCCAGGAGAGAGCTGGTGCAGCCC 182
QY 342 PheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyThrArgAspGly 361
DB 183 TTCAGCTCGCTGTTCCGGAAGTGGAGTACATCGCCAGGCGCGGTGGACCCGGGATGGC 242
QY 362 LysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 381
DB 243 AAATACGCCCTGGGCCATGTTCTCGACCGGCCCGCCAGCAGTGGCTCCAGCTCGTCTCCTC 302
QY 382 ProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAlaArg 401
DB 303 CCCCAGGCCCTGTTATCCCGCAGCAGAGAAATGAGGAGCAGCGGCTAGGCTCTGCCAGA 362
QY 402 AlaValProArgAsnValGlnProTyrValValTyrGluValThrAsnValTrpIle 421
DB 363 GCTGTCCCCAGGAATGTCCAGCCGTATGTGTGTACGAGGAGGTCCACACGCTCTGGATC 422
QY 422 AsnValHisAspIlePheTyrProPheProGlnSerGluGlyGluAspGluLeuCysPhe 441
DB 423 AATGTTTCATGACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGGAGCTCTGCTTT 482
QY 442 LeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeu 461
DB 483 CTCGGGCCCAATGAATGCAAGACGGGCTTCTGCCATTTGTACAAAGTCACCGCGGTTTA 542
QY 462 LysSerGlnGlyTyrAspTrpSerGluProPheSerProGlyGluAspGluPheLysCys 481
DB 543 AAATCCAGGGCTACGATTGGAGTGAGCCCTTCAGCCCCGGGGAGAGATGAATTAAGTGC 602
QY 482 ProIleLysGluGluIleAlaLeuThrSerGlyGluTrp-GluValLeuAlaArgHisG1 501
DB 603 CCATTAAGGAAGAGATTGCTCTGACCAGCGGTGAATGGGAGGNTTTGGCGGAGCACGG 662
QY 501 ySerLysIleTrpValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAspTh 521
DB 663 CTTCAAGATCTGGGTCAATGAAGAGACCAAGCTGGTGTACTTTTCAGGGCACCAGGACAC 722
QY 521 rProLeuGluHisHisLeuTyrValValSerTyrGluAlaAlaGlyGluIleValArgLe 541
DB 723 GCGCTGGAGCACCACCTCTACGTGGTCACTATGAGCGCGCGGAGAACCTAGGCCT 782
QY 541 uThrThr-ProGlyPheSerHisSerCysSerMet-SerGlnAsnProAspMetPheVal 560
DB 783 CACCACCTCCGGGTCTTCCCTAAGCTGCTCCATGAAGCCAAAAATTC----- 830
QY 561 SerHisTyrSerSer-----ValSerThrProPro 570
DB 831 CCACATGGTTCCGAGGCACTTAGGGGATCTCTTAACACCCCCCC 875
```

Search completed: December 12, 2002, 13:53:23
Job time : 1966 secs

THIS PAGE BLANK (uspto)